GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 12:08:22 ; Search time 14 Seconds
(without alignments)
918.404 Million cell updates/sec

Title: US-09-914-152-1
Perfect score: 1657
Sequence: 1 MAFPKMRLMYICLLVLGALC.....TLLDYWQALENSRGEDCPPV 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1657	100.0	310	1	B3G5_HUMAN	Q9y2c3 h beta-1,3-
2	1592	96.1	300	1	B3G5_GORGO	Q9n293 g beta-1,3-
3	1570	94.7	301	1	B3G5_PANPA	Q9n294 p beta-1,3-
4	1552	93.7	297	1	B3G5_PANTR	Q9n295 p beta-1,3-
5	1206	72.8	308	1	B3G5_MOUSE	Q9ji67 m beta-1,3-
6	416	25.1	372	1	B3G8_HUMAN	Q9y2a9 h probable
7	360	21.7	398	1	B3G7_HUMAN	Q9ny97 h beta-1,3-
8	105.5	6.4	1049	1	VP39_YEAST	Q07468 saccharomyc
9	93.5	5.6	336	1	MTHB_HAEPH	P50193 haemophilus
10	90.5	5.5	1314	1	SS22_YEAST	P25390 saccharomyc
11	89.5	5.4	786	1	TLR1_HUMAN	Q15399 homo sapien
12	88.5	5.3	579	1	YHV0_YEAST	P38848 saccharomyc
13	86.5	5.2	516	1	YVNF_AZOCH	P24423 azotobacter
14	86.5	5.2	711	1	ETF2_MYXVL	Q9q8k4 myxoma viru
15	85.5	5.2	632	1	Y242_MYCPN	P75440 mycoplasma
16	85	5.1	1073	1	HSER_PIG	P55204 sus scrofa
17	84.5	5.1	454	1	YE62_HAEIN	P45217 haemophilus
18	84	5.1	306	1	YFV9_SCHPO	Q9p7q6 schizosacch
19	83	5.0	275	1	ISPH_THEMA	Q9x1f7 thermotoga
20	82.5	5.0	525	1	DIMH_CAEEL	O17397 caenorhabdi
21	82.5	5.0	713	1	PMIP_HUMAN	Q99797 homo sapien
22	82	4.9	576	1	HXT4_YEAST	P32467 saccharomyc
23	81.5	4.9	359	1	UNG_YEAST	P12887 saccharomyc
24	81.5	4.9	1042	1	SYI_BORBU	O51773 borrelia bu
25	81	4.9	290	1	LEP_HELPJ	Q9zlq5 helicobacte
26	81	4.9	397	1	O22A_DROME	P81909 drosophila
27	81	4.9	399	1	TRPB_BUCAP	P42391 buchnera ap
28	81	4.9	1157	1	N133_YEAST	P36161 saccharomyc
29	80.5	4.9	290	1	LEP_HELPY	O25300 helicobacte
30	80.5	4.9	711	1	ETF2_SFVKA	Q9q8y2 shope fibro
31	80.5	4.9	3386	1	POLG_DEN4	P09866 d genome po
32	79.5	4.8	612	1	BIR3_MOUSE	Q62210 mus musculu
33	79.5	4.8	691	1	YIO4_YEAST	P40460 saccharomyc
34	78.5	4.7	757	1	METE_PASMU	P57843 pasteurella
35	78.5	4.7	1002	1	YA37_ANASP	P58612 anabaena sp
36	78.5	4.7	4377	1	ANK3_HUMAN	Q12955 homo sapien
37	78	4.7	469	1	LEU2_BUCAP	O85065 buchnera ap
38	78	4.7	519	1	YG31_METJA	Q59025 methanococc
39	78	4.7	618	1	BIR3_HUMAN	Q13490 homo sapien
40	78	4.7	741	1	GNT5_HUMAN	Q09328 homo sapien
41	78	4.7	908	1	DPO1_BORBU	O51498 borrelia bu
42	77.5	4.7	271	1	Y398_METJA	Q57841 methanococc
43	77.5	4.7	466	1	ARLY_CLOPE	Q8xmj8 clostridium
44	77.5	4.7	710	1	PMIP_RAT	Q01992 rattus norv
45	77	4.6	246	1	Y402_BUCAI	P57482 buchnera ap

ALIGNMENTS

RESULT 1
B3G5_HUMAN
ID B3G5_HUMAN STANDARD; PRT; 310 AA.
AC Q9Y2C3; Q9NY96; Q9P1X7; Q9P1X6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5).
GN B3GALT5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal adenocarcinoma;
RX MEDLINE=99230269; PubMed=10212226;
RA Isshiki S., Togayachi A., Kudo T., Nishihara S., Watanabe M.,
RA Kubota T., Kitajima M., Shiraishi N., Sasaki K., Andoh T.,
RA Narimatsu H.;
RT "Cloning, expression, and characterization of a novel UDP-
RT galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
RT (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal
RT and pancreatic epithelia and tumor cells derived therefrom.";
RL J. Biol. Chem. 274:12499-12507(1999). April 99.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99337698; PubMed=10406968;
RA Zhou D., Berger E.G., Hennet T.;
RT "Molecular cloning of a human UDP-galactose:GlcNAcbeta1,3GalNAc
RT beta1, 3 galactosyltransferase gene encoding an O-linked
RT core3-elongation enzyme.";
RL Eur. J. Biochem. 263:571-576(1999). July 99.
RN [3]
RP SEQUENCE FROM N.A.
RA Amado M., Carneiro F., Clausen H.;
RT "Cloning and expression of two beta-1,3-galactosyltransferases:
RT beta3gal-T5 and beta3gal-T6.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,

```
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [5]
RP   SEQUENCE OF 1-298 FROM N.A.
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   REVIEW.
RX   MEDLINE=20047730; PubMed=10580128;
RA   Amado M., Almeida R., Schwientek T., Clausen H.;
RT   "Identification and characterization of large galactosyltransferase
RT   gene families: galactosyltransferases for all functions.";
RL   Biochim. Biophys. Acta 1473:35-53(1999).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR.
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN STOMACH, JEJUNUM, COLON,
CC       PANCREAS, SMALL INTESTINE, TESTIS AND GASTROINTESTINAL AND
CC       PANCREATIC CANCER CELL LINES. HARDLY DETECTED IN LUNG, LIVER,
CC       ADRENAL GLAND AND PERIPHERAL BLOOD LEUKOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB020337; BAA77664.1; -.
DR   EMBL; AF145784; AAF07880.1; -.
DR   EMBL; AJ006078; CAB91547.1; -.
DR   EMBL; AL163280; CAB90446.1; -.
DR   EMBL; AB041412; BAA94497.1; -.
DR   EMBL; AB041413; BAA94498.1; -.
DR   Genew; HGNC:920; B3GALT5.
DR   MIM; 604066; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN       1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     8      28      SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                               (POTENTIAL).
FT   DOMAIN       29     310     LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     130    130     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     174    174     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     231    231     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     23     23      F -> Y (IN REF. 3).
FT   CONFLICT     26     26      Y -> N (IN REF. 3).
FT   CONFLICT     42     42      G -> V (IN REF. 5; BAA94497).
FT   CONFLICT     85     85      M -> T (IN REF. 3 AND 5; BAA94497).
FT   CONFLICT     114    114     Q -> R (IN REF. 5; BAA94497).
FT   CONFLICT     234    234     K -> E (IN REF. 5; BAA94497).
FT   CONFLICT     277    277     L -> R (IN REF. 5; BAA94497).
FT   CONFLICT     293    297     LDYWQ -> WTTGR (IN REF. 5; BAA94498).
SQ   SEQUENCE   310 AA;  36189 MW;  4DD7A19E3E648AA9 CRC64;

  Query Match             100.0%;  Score 1657;  DB 1;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-146;
  Matches  310;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60

Qy      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120

Qy     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180

Qy     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240

Qy     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300

Qy     301 NSRGEDCPPV 310
           ||||||||||
Db     301 NSRGEDCPPV 310


RESULT 2
B3G5_GORGO
ID   B3G5_GORGO     STANDARD;      PRT;   300 AA.
AC   Q9N293;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE   (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN   B3GALT5.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX   NCBI_TaxID=9595;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR (BY SIMILARITY).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB041416; BAA94501.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN       1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     8      28      SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                               (POTENTIAL).
FT   DOMAIN       29     >300    LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     130    130     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     174    174     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     231    231     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER      300    300
SQ   SEQUENCE   300 AA;  35104 MW;  1AA11692ED9F06FA CRC64;

  Query Match             96.1%;  Score 1592;  DB 1;  Length 300;
  Best Local Similarity   99.7%;  Pred. No. 6.5e-140;
  Matches  299;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60

Qy      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      61 LLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120

Qy     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180

Qy     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240

Qy     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300


RESULT 3
B3G5_PANPA
ID   B3G5_PANPA     STANDARD;      PRT;   301 AA.
AC   Q9N294;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE   (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN   B3GALT5.
OS   Pan paniscus (Pygmy chimpanzee) (Bonobo).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9597;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR (BY SIMILARITY).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB041415; BAA94500.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29   >301       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    231    231       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     301    301
SQ   SEQUENCE   301 AA;  35222 MW;  E04CC70F41CB29FD CRC64;
```

```
  Query Match           94.7%;  Score 1570;  DB 1;  Length 301;
  Best Local Similarity  98.0%;  Pred. No. 7.1e-138;
  Matches  295;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60
           ||||||||||:|||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1 MAFPKMRLMYVCLLVLGALCLYFSMYSLNLFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60

Qy      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      61 LLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120

Qy     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180
           ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QKDFLDGYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180

Qy     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSESVPYIK 240

Qy     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCRFRRIVACHFIKPRTLLDYWQALE 300

Qy     301 N 301
           |
Db     301 N 301


RESULT 4
B3G5_PANTR
ID   B3G5_PANTR     STANDARD;      PRT;   297 AA.
AC   Q9N295;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE   (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN   B3GALT5.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR (BY SIMILARITY).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB041414; BAA94499.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
```

```
FT   DOMAIN       29   >297       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    231    231       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     297    297
SQ   SEQUENCE   297 AA;  34850 MW;  BB7963250A837A28 CRC64;

  Query Match             93.7%;  Score 1552;  DB 1;  Length 297;
  Best Local Similarity   97.6%;  Pred. No. 3.2e-136;
  Matches  290;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60
           ||||||||||:|||||||||:|||||||| |||||||||||||||||||||||||||||
Db       1 MAFPKMRLMYVCLLVLGALCVYFSMYSLNLFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60

Qy      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120
           |||||||:|||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      61 LLVTSSHRQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120

Qy     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180

Qy     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
           ||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSESVPYIK 240

Qy     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQ 297
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCRFRRIVACHFIKPRTLLDYWQ 297


RESULT 5
B3G5_MOUSE
ID   B3G5_MOUSE     STANDARD;      PRT;   308 AA.
AC   Q9JI67;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE   (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 5) (Beta-3-Gx-T5) (Stage-specific embryonic
DE   antigen-3 synthase) (SSEA-3 synthase).
GN   B3GALT5 OR B3GT5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RX   MEDLINE=20390006; PubMed=10837462;
RA   Zhou D., Henion T.R., Jungalwala F.B., Berger E.G., Hennet T.;
RT   "The beta1,3-galactosyltransferase beta3GalT-V is a stage-specific
RT   embryonic antigen-3 (SSEA-3) synthase.";
RL   J. Biol. Chem. 275:22631-22634(2000).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR. ALSO CATALYZES THE TRANSFER OF GAL TO THE
CC       TERMINAL GALNAC UNIT OF THE GLOBOSIDE GB4, THEREBY SYNTHESIZING
CC       THE GLYCOLIPID GB5, ALSO KNOWN AS THE STAGE-SPECIFIC EMBRYONIC
CC       ANTIGEN-3 (SSEA-3).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF254738; AAF86241.1; -.
DR   MGD; MGI:2136878; B3galt5.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     25       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       26    308       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    172    172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   308 AA;  35964 MW;  789073A5178825B1 CRC64;

  Query Match             72.8%;  Score 1206;  DB 1;  Length 308;
  Best Local Similarity   71.7%;  Pred. No. 3.4e-104;
  Matches  223;  Conservative   40;  Mismatches   44;  Indels    4;  Gaps    2;

Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKD-GNFLKLPDTDCRQTPPFL 59
           ||  | ||:|  :|::|||||||| |   |:|  ||:||  | ||::|| ||:| ||||
Db       1 MAHMKTRLVYASILMMGALCLYFSMDS---FRELPFVFKKSHGKFLQIPDIDCKQKPPFL 57

Qy      60 VLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDI 119
           |||||||||||| |||||:|||:|  |:|:|::|||||||: |  |       ||::| ||
Db      58 VLLVTSSHKQLAARMAIRKTWGRETSVQGQQVRTFFLLGTSDSTEEMDATTLESEQHRDI 117

Qy     120 IQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRF 179
           ||||| | |:||||||||||:|||: |||| |:|||||||||:|| ||||||||||:||||
Db     118 IQKDFKDAYFNLTLKTMMGMEWVYHFCPQTAYVMKTDSDMFVNVGYLTELLLKKNKTTRF 177

Qy     180 FTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYI 239
           |||::| ::|||||  |:||||||  |||||||||||||||||||| ||| ||||||:|||:|
Db     178 FTGYIKPHDFPIRQKFNKWFVSKFEYPWDRYPPFCSGTGYVFSSDVAIQVYNVSESVPFI 237

Qy     240 KLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 299
           |||||||||||  :| || ||||:: |||||||||||| |::|||||:||: || |||||
Db     238 KLEDVFVGLCLAKLKIRPEELHTKQTFFPGGLRFSVCRFQKIVACHFMKPQDLLTYWQAL 297

Qy     300 ENSRGEDCPPV 310
           |||: :||| |
Db     298 ENSKEQDCPAV 308


RESULT 6
B3G8_HUMAN
ID   B3G8_HUMAN     STANDARD;      PRT;   372 AA.
AC   Q9Y2A9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable beta-1,3-galactosyltransferase 8 (EC 2.4.1.-) (Beta-1,3-
DE   GalTase 8) (Beta3Gal-T8) (b3Gal-T8) (UDP-galactose:beta-N-
DE   acetylglucosamine beta-1,3-galactosyltransferase 8) (UDP-Gal:beta-
DE   GlcNAc beta-1,3-galactosyltransferase 8) (Beta-3-Gx-T8).
GN   B3GALT8 OR TMEM3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99173880; PubMed=10072769;
RA   Yokoyama-Kobayashi M., Yamaguchi T., Sekine S., Kato S.;
RT   "Selection of cDNAs encoding putative type II membrane proteins on the
RT   cell surface from a human full-length cDNA bank.";
RL   Gene 228:161-167(1999).
RN   [2]
RP   REVIEW.
```

```
RX   MEDLINE=20047730; PubMed=10580128;
RA   Amado M., Almeida R., Schwientek T., Clausen H.;
RT   "Identification and characterization of large galactosyltransferase
RT   gene families: galactosyltransferases for all functions.";
RL   Biochim. Biophys. Acta 1473:35-53(1999).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015630; BAA76497.1; -.
DR   Genew; HGNC:13528; B3GNT3.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     11     31       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       32    372       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    202    202       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    362    362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   372 AA;  42515 MW;  A6E3FE88AFCCFCC2 CRC64;

  Query Match             25.1%;  Score 416;  DB 1;  Length 372;
  Best Local Similarity   34.0%;  Pred. No. 5.3e-31;
  Matches  100;  Conservative   57;  Mismatches  109;  Indels   28;  Gaps    8;

Qy      26 YSLNPFKEQSF-VYKKDGNFLKL---PDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWG 81
           ::  |   |:| :|:   :|  |   | : | | | ||:|:: ||       |  :|:|||
Db      73 FATQPQHVQNFLLYRHCRHFPLLQDVPPSKCAQ-PVFLLLVIKSSPSNYVRRELLRRTWG 131

Qy      82 KERMVKGKQLKTFFLLGTTSSAAETKEVDQ----ESQRHGDIIQKDFLDVYYNLTLKTMM 137
           :|| |:| ||:  ||:|| |:  | ::|::    |:| ||||:| || | ::|||| ::
Db     132 RERKVRGLQLRLLFLVGTASNPHEARKVNRLLELEAQTHGDILQWDFHDSFFNLTLKQVL 191

Qy     138 GIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSK 197
            ::|    |  |:||:  | |:| : | :   |   :     | | |  | |||  :||
Db     192 FLQWQETRCANASFVLNGDDDVFAHTDNMVFYLQDHDPGRHLFVGQLIQNVGPIRAFWSK 251

Qy     198 WFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRL 257
           ::| :     :||||:| | |:: |   |: :   :  :    ::|||:|:|||     |
Db     252 YYVPEVVTQNERYPPYCGGGGFLLSRFTAAALRRAAHVLDIFPIDDVFLGMCLE-----L 306

Qy     258 EELHSQPTFFPG----GLR--------FSVCLFRRIVACHFIKPRTLLDYWQAL 299
           | |  :|    |    |:|         | | :| ::  |   |  :|  | ||
Db     307 EGL--KPASHSGIRTSGVRAPSQHLSSFDPCFYRDLLLVHRFLPYEMLLMWDAL 358


RESULT 7
B3G7_HUMAN
ID   B3G7_HUMAN     STANDARD;      PRT;   398 AA.
AC   Q9NY97; Q9NQR0; Q9NUT9; Q9NQQ9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 7 (EC 2.4.1.-) (Beta-1,3-GalTase 7)
DE   (Beta3Gal-T7) (b3Gal-T7) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 7) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 7) (Beta-3-Gx-T7).
GN   B3GALT7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Amado M., Carneiro F., Clausen H.;
RT   "Cloning and expression of two beta-1,3-galactosyltransferases:
RT   beta3gal-T5 and beta3gal-T6.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Urinary bladder;
RA   Gromova I., Gromov P., Celis J.E.;
RT   "A novel member of beta-1,3-galactosyltransferase family is down
RT   regulated during bladder TCC progression.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 25-398 FROM N.A.
RC   TISSUE=Placenta;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA   Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA   Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA   Masuho Y., Kanehori K.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVIEW.
RX   MEDLINE=20047730; PubMed=10580128;
RA   Amado M., Almeida R., Schwientek T., Clausen H.;
RT   "Identification and characterization of large galactosyltransferase
RT   gene families: galactosyltransferases for all functions.";
RL   Biochim. Biophys. Acta 1473:35-53(1999).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   -!- CAUTION: WAS INDICATED AS B3GAL-T6 IN SUBMITTED DNA ENTRIES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ006077; CAB91546.1; -.
DR   EMBL; AF288208; AAF97253.1; -.
DR   EMBL; AF288209; AAF97254.1; -.
DR   EMBL; AK002009; BAA92031.1; ALT_INIT.
DR   Genew; HGNC:15629; B3GNT1.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family; Alternative splicing.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    398       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    220    220       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     12       MSVGRRRIKLLL -> MVSRSLV (IN ISOFORM 2).
FT   CONFLICT     12     12       MISSING (IN REF. 2).
SQ   SEQUENCE   398 AA;  46135 MW;  115B829FEB51D2F8 CRC64;

  Query Match             21.7%;  Score 360;  DB 1;  Length 398;
  Best Local Similarity   29.1%;  Pred. No. 8.7e-26;
  Matches   82;  Conservative   66;  Mismatches  120;  Indels   14;  Gaps    7;
```

```
Qy      31 FKEQSFVYKKDGNFLKL---PDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVK 87
           ||:   :| :  |:  |   ||   ::  |||:| : |     | | |||::||:|
Db     117 FKD-FLLYLRCRNYSLLIDQPDKCAKK--PFLLLAIKSLTPHFARRQAIRESWGQESNAG 173

Qy      88 GKQLKTFFLLGTTSSAAETKEVDQ----ESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVH 143
             : :   |||| |        ::      ||::| ||:  :: | ::||:|| :: : ||
Db     174 NQTVVRVFLLGQTPPEDNHPDLSDMLKFESEKHQDILMWNYRDTFFNLSLKEVLFLRWVS 233

Qy     144 RFCPQAAFVMKTDSDMFINVDYLTELL--LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVS 201
             ||   || | | |:|:|  ::   |  | | :    | | :  |  | |    |:::
Db     234 TSCPDTEFVFKGDDDVFVNTHHILNYLNSLSKTKAKDLFIGDVIHNAGPHRDKKLKYYIP 293

Qy     202 KSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELH 261
           :  |    |||:  | |:::|| :| ::|:::  |    ::|||: |:||::| :  |:
Db     294 EVVYS-GLYPPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLQKLGLVPEKHK 352

Qy     262 SQPTF-FPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALENS 302
             ||      : ::| :  ::  |  ||: ::| |  |:::
Db     353 GFRTFDIEEKNKNNICSYVDLMLVHSRKPQEMIDIWSQLQSA 394


RESULT 8
VP39_YEAST
ID   VP39_YEAST     STANDARD;      PRT;  1049 AA.
AC   Q07468;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Vacuolar assembly protein VPS39 (Vacuolar morphogenesis protein VAM6).
GN   VPS39 OR VAM6 OR YDL077C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97269041; PubMed=9111041;
RA   Nakamura N., Hirata A., Ohsumi Y., Wada Y.;
RT   "Vam2/Vps41p and Vam6/Vps39p are components of a protein complex on
RT   the vacuolar membranes and involved in the vacuolar assembly in the
RT   yeast Saccharomyces cerevisiae.";
RL   J. Biol. Chem. 272:11344-11349(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wambutt R., Wedler H., Wedler E., Scharfe M.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR VACUOLAR ASSEMBLY.
CC   -!- SUBUNIT: INTERACTS WITH VPS41/VAM2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74125; CAA98643.1; -.
DR   EMBL; D83058; BAA11758.1; -.
DR   SGD; S0002235; VAM6.
SQ   SEQUENCE   1049 AA;  122881 MW;  EE8D94E7E2F5F078 CRC64;

  Query Match             6.4%;  Score 105.5;  DB 1;  Length 1049;
  Best Local Similarity   24.4%;  Pred. No. 0.1;
  Matches   44;  Conservative   32;  Mismatches   53;  Indels   51;  Gaps    9;

Qy      37 VYKKDGNFLKL----PD-------------------------TDCRQTPPFLVLLVTS 65
           || :||| |||    ||                         |::|   |:|| |:
Db      38 VYSRDGNTLKLFQVYPDLLQNAKNDPLPPVIENFYFANELSTIFAQCKET---LILLSTT 94

Qy      66 S---HKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQ- 121
           :   : :: :|  |   |  ||  | |: |  :|: :| : |: : :  | : : ::::
Db      95 NLHEYDRIIDRRGINHCWLFERSHKNKEEKNTYLIYSTINTAKMRVLIWEGRTYKNMMEA 154

Qy     122 -----KDFLDVYY----NLTLKTMMGI-EWVHRFCPQAAFVMKTDSDMFIN--VDYLTEL 169
                |: :   |     :|| | :||  | :   |    : || : |    :  ||||
Db     155 SLSYRKETIRSIYPGETGITLATDLGIYHWPYN-KPSLIRIEKTVKNKFPKDMISALTEL 213


RESULT 9
MTHB_HAEPH
ID   MTHB_HAEPH     STANDARD;      PRT;   336 AA.
AC   P50193;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Modification methylase HphIB (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase HphIB) (M.HphIB) (M.HphI(A)).
GN   HPHIBM OR HPHIMA.
OS   Haemophilus parahaemolyticus.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=735;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49700;
RX   MEDLINE=96313243; PubMed=8759008;
RA   Lubys A., Lubiene J., Kulakauskkas S., Stankevicius K.,
RA   Timinskas A., Janulaitis A.;
RT   "Cloning and analysis of the genes encoding the type IIS restriction-
RT   modification system HphI from Haemophilus parahaemolyticus.";
RL   Nucleic Acids Res. 24:2760-2766(1996).
CC   -!- FUNCTION: THIS METHYLASE PROTECTS THE DNA FROM CLEAVAGE BY THE
CC       HPHI ENDONUCLEASE.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85374; CAA59691.1; -.
DR   REBASE; 3661; M.HphIB.
DR   InterPro; IPR002294; D12N6_mtfrase.
DR   InterPro; IPR002052; N6_Mtase.
DR   Pfam; PF02086; MethyltransfD12; 1.
DR   PRINTS; PR00505; D12N6MTFRASE.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Transferase; Methyltransferase; Restriction system.
SQ   SEQUENCE   336 AA;  39609 MW;  3BD144C7772C4FDF CRC64;

  Query Match             5.6%;  Score 93.5;  DB 1;  Length 336;
  Best Local Similarity   22.3%;  Pred. No. 0.32;
  Matches   39;  Conservative   27;  Mismatches   58;  Indels   51;  Gaps    7;

Qy     112 ESQRHG-DIIQKDFLDVYYNLTLK---------TMMGIEWVHRFCPQAAFVMKTDSDMFI 161
           |::: |  :|  | |:: | | |          |   ::::    |:: |: |  ||:|
Db      46 EAKKRGYQVIANDILNINYQLALALIVNNQEILTACDVDFIFSNPPKSGFMTKNYSDVFF 105

Qy     162 ------NVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQ-PFSKWFVS--------KSEYP 206
                 :| :   :|| | | :    |  :    ||: |:|:: :|          ||
Db     106 FKEECRELDAIRANILKLNNTYKQALAFALMRRAMIRKMPYSRFTISWEKVKQLRDEEYS 165

Qy     207 WDRY----------------PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVF 245
           : :|                |   ||:   | ||          ||||
Db     166 YSKYGRRRAYHNQSFEFHFRENLNSYNQAVFNNGNIHQAYN----------EDVF 210


RESULT 10
SS22_YEAST
ID   SS22_YEAST     STANDARD;      PRT;  1314 AA.
AC   P25390;
```

```
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein kinase SSK22 (EC 2.7.-.-).
GN   SSK22 OR YCR073C OR YCR73C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
RA   Sanz E.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95350642; PubMed=7624781;
RA   Maeda T., Takekawa M., Saito H.;
RT   "Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-
RT   containing osmosensor.";
RL   Science 269:554-558(1995).
CC   -!- FUNCTION: KINASE INVOLVED IN A SIGNAL TRANSDUCTION PATHWAY THAT IS
CC       ACTIVATED BY CHANGES IN THE OSMOLARITY OF THE EXTRACELLULAR
CC       ENVIRONMENT. ACTIVATES THE PBS2 MAP KINASE KINASE BY
CC       PHOSPHORYLATION.
CC   -!- SUBUNIT: INTERACTS WITH BY SSK1.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MAP KINASE KINASE KINASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; CAA42271.1; -.
DR   PIR; S19488; S19488.
DR   HSSP; P24941; 1B38.
DR   SGD; S0000669; SSK22.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN     1034   1314       PROTEIN KINASE.
FT   NP_BIND    1040   1048       ATP (BY SIMILARITY).
FT   BINDING    1063   1063       ATP (BY SIMILARITY).
FT   ACT_SITE   1158   1158       BY SIMILARITY.
SQ   SEQUENCE   1314 AA;  150654 MW;  8040A31FAAD56E14 CRC64;

  Query Match             5.5%;  Score 90.5;  DB 1;  Length 1314;
  Best Local Similarity   19.2%;  Pred. No. 3.3;
  Matches   61;  Conservative   58;  Mismatches  127;  Indels   71;  Gaps   14;

Qy      33 EQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERM----AIRQTW-GKERMVK 87
           |||  | :| :    :    | |   :| : :   | : |    ::   |   ::| |
Db     209 EQSLTYLRDSSDSVFEEIMKFQIPQGKILSLDALEAILQDLMNRYHSVVSYWPNLKKMYK 268

Qy      88 GKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 147
            | :       |::|  |  :|        :   :|:   ||||:     :|::|| |
Db     269 DKPI--------TNTAEFTARID---------VMNSWLNFKTNLTLRRQELDDWINRFSP 311

Qy     148 QAAFVMKTDS---------DMFINVDYLTELLLKKNRTTRFFTG----------FLKLN 187
               : :|:             :  | | |:|:      |              |||:
Db     312 ----ISSSDNCQEDFDGVPQWNCKMKILAEQLMKEKNIESIFQKKIFYPLSPWMFKLKLH 367

Qy     188 EFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQV-YNVSKSVPYIKLEDVFV 246
               |:  :|  :   :||::|     :   |:    : ::: |      | : : |  :

Db     368 FIVYRETLTKMNI---KYPYERLRSLLAFPVYLIKEVILTRLSYARKLKNPTMMMIDQMI 424

Qy     247 G--LCLERLNIRLE----------ELHSQPTF---FPGGLRFSVCLFR-RIVAC---H 285
               ||:::|:            ::   |||       :|:   |   :::     :
Db     425 DDFNAFIRLSVQLKYTLTKYCSNLPFDVDFDPTFENTVIEAIRYLFFLLNLKLIDSSKQN 484

Qy     286 FIKPRTLLDYWQALENS 302
           |  |  || ||  |:|:
Db     485 FKAPDLLLKYWDHLKNT 501


RESULT 11
TLR1_HUMAN
ID   TLR1_HUMAN     STANDARD;      PRT;   786 AA.
AC   Q15399; O15452; Q9UG90;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Toll-like receptor 1 precursor (Toll/interleukin-1 receptor-like)
DE   (TIL).
GN   TLR1 OR KIAA0012.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Erythroleukemia;
RX   MEDLINE=98118556; PubMed=9435236;
RA   Rock F.L., Hardiman G., Timans J.C., Kastelein R.A., Bazan J.F.;
RT   "A family of human receptors structurally related to Drosophila
RT   Toll.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:588-593(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051387; PubMed=7584026;
RA   Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA   Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. 1.
RT   The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT   analysis of randomly sampled cDNA clones from human immature myeloid
RT   cell line KG-1.";
RL   DNA Res. 1:27-35(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21154917; PubMed=11230166;
RA   Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA   Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA   Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA   Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA   Wambutt R., Korn B., Klein M., Poustka A.;
RT   "Towards a catalog of human genes and proteins: sequencing and
RT   analysis of 500 novel complete protein coding human cDNAs.";
RL   Genome Res. 11:422-435(2001).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF TIR DOMAIN.
RX   MEDLINE=20531768; PubMed=11081518;
RA   Xu Y., Tao X., Shen B., Horng T., Medzhitov R., Manley J.L., Tong L.;
RT   "Structural basis for signal transduction by the Toll/interleukin-1
RT   receptor domains.";
RL   Nature 408:111-115(2000).
CC   -!- FUNCTION: Participates in the innate immune response to microbial
CC       agents. Cooperates with TLR2 and modulates the response to
CC       microbial constituents. Acts via MyD88 and TRAF6, leading to NF-
CC       kappa-B activation, cytokine secretion and the inflammatory
CC       response (By similarity).
CC   -!- SUBUNIT: Binds TLR2 via their respective extracellular domains.
CC       Binds MyD88 via their respective TIR domains (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Plasma membrane and
CC       phagosomes (By similarity).
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in spleen, ovary,
```

```
CC       peripheral blood leukocytes, thymus and small intestine.
CC   -!- SIMILARITY: BELONGS TO THE TOLL-LIKE RECEPTOR FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TIR DOMAIN.
CC   -!- SIMILARITY: CONTAINS 8 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U88540; AAC34137.1; -.
DR   EMBL; D13637; BAA02801.1; -.
DR   EMBL; AL050262; CAB43364.1; -.
DR   Genew; HGNC:11847; TLR1.
DR   MIM; 601194; -.
DR   PDB; 1FYW; 22-NOV-00.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR000157; TIR_domain.
DR   Pfam; PF00560; LRR; 7.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01582; TIR; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00369; LRR_TYP; 4.
DR   SMART; SM00255; TIR; 1.
DR   PROSITE; PS50104; TIR; 1.
KW   Receptor; Immune response; Inflammatory response; Signal;
KW   Transmembrane; Repeat; Leucine-rich repeat; Glycoprotein;
KW   3D-structure.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    786       TOLL-LIKE RECEPTOR 1.
FT   DOMAIN      23    580       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   581    601       POTENTIAL.
FT   DOMAIN     602    786       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      68     91       LRR 1.
FT   REPEAT      93    113       LRR 2.
FT   REPEAT     114    137       LRR 3.
FT   REPEAT     139    164       LRR 4.
FT   REPEAT     371    395       LRR 5.
FT   REPEAT     397    423       LRR 6.
FT   REPEAT     444    467       LRR 7.
FT   REPEAT     469    491       LRR 8.
FT   DOMAIN     635    779       TIR.
FT   CARBOHYD    51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   137    137       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   248    248       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   429    429       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    80     80       R -> T (IN REF. 2).
FT   CONFLICT   182    182       E -> G (IN REF. 3).
FT   CONFLICT   228    228       N -> S (IN REF. 3).
FT   CONFLICT   248    248       S -> N (IN REF. 2).
FT   CONFLICT   276    276       F -> S (IN REF. 3).
FT   CONFLICT   602    602       I -> S (IN REF. 2).
SQ   SEQUENCE   786 AA;  90290 MW;  1406C60FB5847F58 CRC64;

  Query Match             5.4%;  Score 89.5;  DB 1;  Length 786;
  Best Local Similarity   22.4%;  Pred. No. 2.2;
  Matches   49;  Conservative   38;  Mismatches   79;  Indels   53;  Gaps   12;

Qy     121 QKDFLDVYYNLTLK--TMMGIEWVHRFCPQ--AAFV-------MKTDSDMFINVDYLTE- 168
           |    ::: |: :|  |: |   ||  ||   : |:       : ||: :| |  :|||
Db     316 QSYIYEIFSNMNIKNFTVSGTRMVHMLCPSKISPFLHLDFSNNLLTDT-VFENCGHLTEL 374

Qy     169 --LLLKKNRTTRFFTGFLKLNEFPIR-QPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDV 225
             |:|: |:       |: |   : :   :  :|::   :|     || |  : | ::

Db     375 ETLILQMNQLKE----LSKIAEMTTQMKSLQQLDISQNSVSYDEKKGDCSWTKSLLSLNM 430

Qy     226 ASQVY-------------------NVSKSVP--YIKLEDVFVGLCLERLNIRLEELHSQP 264
           :| :                     |  ||:|   :|||        |: ||:   |
Db     431 SSNILTDTIFRCLPPRIKVLDLHSNKIKSIPKQVVKLE------ALQELNVAFNSL---- 480

Qy     265 TFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALENSR 303
           |  ||   ||      ::  |        |::|: :  |
Db     481 TDLPGCGSFS--SLSVLIIDHNSVSHPSADFFQSCQKMR 517


RESULT 12
YHV0_YEAST
ID   YHV0_YEAST     STANDARD;      PRT;   579 AA.
AC   P38848;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Hypothetical 66.1 kDa protein in IMP3-SPO12 intergenic region.
GN   YHR150W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U10397; AAB68980.1; -.
DR   PIR; S46757; S46757.
DR   SGD; S0001193; YHR150W.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM    237    253       POTENTIAL.
FT   TRANSMEM    397    421       POTENTIAL.
SQ   SEQUENCE   579 AA;  66148 MW;  2A6A73999C54C19E CRC64;

  Query Match             5.3%;  Score 88.5;  DB 1;  Length 579;
  Best Local Similarity   27.4%;  Pred. No. 1.8;
  Matches   37;  Conservative   22;  Mismatches   51;  Indels   25;  Gaps    7;

Qy      81 GKERMVKGKQL--KTFF--LLGTTSSAAETKEVDQESQRHGDIIQ--KDFLDV--YYNLT 132
           ||     |:|   |:|   |        |  |:|: ::  |:: |  |  :::    |:|
Db     304 GKNEYSYGQQFFSKSFMDTLESRNQEIDEISELDKRTENTGELKQGMKVLINLRDMQNMT 363

Qy     133 LKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIR 192
           |:  || ::      :|: |: |  | |         ::  | |||||| :    |
Db     364 SGTLHVIEAIN------SFLRKSSS--FQN---------EECSTKRFFTGFLLIVFLKIL 406

Qy     193 QPFSKWFVSKSEYPW 207
           || |    | : |
Db     407 SPFVNWSYVCSIFAW 421


RESULT 13
```

```
YVNF_AZOCH
ID   YVNF_AZOCH     STANDARD;      PRT;   516 AA.
AC   P24423;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein in vnfD 5'region.
OS   Azotobacter chroococcum mcd 1.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Azotobacter.
OX   NCBI_TaxID=355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90356423; PubMed=2388847;
RA   Fallik E., Robson R.L.;
RT   "Completed sequence of the region encoding the structural genes for
RT   the vanadium nitrogenase of Azotobacter chroococcum.";
RL   Nucleic Acids Res. 18:4616-4616(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51756; CAA36057.1; -.
DR   PIR; S14694; S14694.
KW   Nitrogen fixation; Hypothetical protein.
SQ   SEQUENCE   516 AA;  58676 MW;  ED7A7F362ED08A14 CRC64;

  Query Match             5.2%;  Score 86.5;  DB 1;  Length 516;
  Best Local Similarity   24.8%;  Pred. No. 2.4;
  Matches   66;  Conservative   30;  Mismatches   97;  Indels   73;  Gaps   12;

Qy      55 TPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQ--- 111
           || ||| |:|       || |   |   |:::    | : | :|  |: |   :
Db     284 TPTFLVDLLT-------ERRAF--TPALEQVMATDALLSAFEVGDISTEALMFQAGYLTI 334

Qy     112 -ESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
            : || |   |  :   | || :|  :    : |: |: | ||:    ::        :||
Db     335 GQVQRIG--AQPRYRLRYPNLEVKASLNEALLLRYLPEPASVMRQTGQLY-------DLL 385

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |               |:||  |     | :   : | |  |     ||      ||  |
Db     386 L--------------ANDFPGLQTLFTAFFASIPHDWYRNNPIARYEGY-----YASVFY 426

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
           :            |  |    |::|||:  |      | :  :| ||   |     |
Db     427 S------------HFAAL---GLDVRLEDATSH-----GRIDMTV-LFNGHVYLFEFKVV 465

Qy     291 TLLDYWQALE-----------NSRGE 305
           |    |||:            :|||
Db     466 ELTPAGQALQQLKDKGYADKYRARGE 491


RESULT 14
ETF2_MYXVL
ID   ETF2_MYXVL     STANDARD;      PRT;   711 AA.
AC   Q9Q8K4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Early transcription factor 82 kDa subunit (VETF large subunit).
GN   M096L.
OS   Myxoma virus (strain Lausanne).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Leporipoxvirus.
OX   NCBI_TaxID=31530;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20032073; PubMed=10562494;
RA   Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.X.,
RA   Macaulay C., Willer D., Evans D., McFadden G.;
RT   "The complete DNA sequence of myxoma virus.";
RL   Virology 264:298-318(1999).
CC   -!- FUNCTION: ACTS WITH RNA POLYMERASE TO INITIATE TRANSCRIPTION FROM
CC       EARLY GENE PROMOTERS. A DNA-DEPENDENT ATPASE ACTIVITY IS
CC       ASSOCIATED WITH VETF.
CC   -!- SUBUNIT: HETERODIMER OF A 70 kDa AND A 82 kDa SUBUNIT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF170726; AAF14984.1; -.
KW   Transcription regulation; Activator.
SQ   SEQUENCE   711 AA;  82687 MW;  F69EC4F63B087E61 CRC64;

  Query Match             5.2%;  Score 86.5;  DB 1;  Length 711;
  Best Local Similarity   20.1%;  Pred. No. 3.6;
  Matches   65;  Conservative   55;  Mismatches  118;  Indels   85;  Gaps   16;

Qy      29 NPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKG 88
           |  : :| :|| |  |: |       |   | | |  |     : :    : |   |:
Db     201 NGSRYKSLLYKSDEYFINLTGNKFILTDERLNLTVWDS-----DGVVTFSSDGDTIMINN 255

Qy      89 KQLKTFFLLGTTSSAAETKEVDQESQR-HGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 147
             :| |  |           ::| : :|  ||:  | ||     :| :  : ||
Db     256 VKLFTSLL----------TDIDLQMERIKGDVTYKIFLST--PITSRIKLNIE------T 297

Qy     148 QAAFVMKTDSDMFINVDYLTELLLKKNRTT----------RFFTGFLKL---NEFPIRQ 193
             ||    ::: :: |    ::| ||  :           ::|| || :   : |   |
Db     298 SFIFVETATNNILLSADKRISIILAKNHISIKVKNYIPNIEKYFT-FLVIAINSMFNNIQ 356

Qy     194 PFSKWFVSKSEYPWDRYPPFCSGTGY-----VFSGDVASQVYNVSKSVPYIKLEDVFVG- 247
           | :   :: | | |    | |          |    :   :  || :      ::||:
Db     357 QASDFTKVETVY-WSR---ICQNTKNKHRKPVIVSSLDEDMEKVSDNFYKSSTKEVFINS 412

Qy     248 -----LCLERLN----------IRLEEL----------HSQPTFFPGGLRFSVCLFRR 280
                ||: ||           ||:::          |::          || |::::
Db     413 SGIMFSCLDPLNKYNYVGFLSIFYRLQKMCIPCCFLKNQSHTE--------TFSSCVYQK 464

Qy     281 IVACHFIKPRTLLDYWQALENSR 303
           :|   | |  :|:: : :  |:
Db     465 EIASDVINP-YILNFGKVVTKSK 486


RESULT 15
Y242_MYCPN
ID   Y242_MYCPN     STANDARD;      PRT;   632 AA.
AC   P75440;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MG242 homolog (F10_orf632o).
GN   MPN338 OR MP498.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000049; AAB96146.1; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     255     275       POTENTIAL.
FT   TRANSMEM     506     526       POTENTIAL.
FT   TRANSMEM     566     586       POTENTIAL.
FT   TRANSMEM     603     623       POTENTIAL.
SQ   SEQUENCE   632 AA;  74335 MW;  344D6771FCDA44A6 CRC64;

  Query Match             5.2%;  Score 85.5;  DB 1;  Length 632;
  Best Local Similarity   18.6%;  Pred. No. 3.9;
  Matches   49;  Conservative   47;  Mismatches   78;  Indels   89;  Gaps   13;

Qy       3 FPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKL-------PDTDCRQT 55
           :  |  : | |:::  | :||:       || : : |   | : |       |:||  :
Db     257 YATMNTLLIMLIIIEELRVYFNS------KEPALILKLLDNKVSLREDPDQNPETDLHE- 309

Qy      56 PPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQR 115
                |:  : |   |:    : | |:|:   ::|              | |::: |::
Db     310 ------LIKFAEKNYLEKEKTSR-WHKKRVKSLEEL-----------LEEIKQINLETKN 351

Qy     116 HG-----DIIQKDFLDVY----------YNLTLKTMMGI----------EWVHRFCPQAA 150
                  :|:: :  :|:          : | |:|: ||           | : |
Db     352 ESLAYPDEIVELELDNVHNFVSTKQVFRHQLDLQTLHGIVINPERYGIGMWSNHFVDWEE 411

Qy     151 FV--------MKTDSDMF-----------INVDYLT-------ELLLKKNRTTRFFTGF 183
           |          :  ||::            :|  |||        |: ||  |:  : :
Db     412 FKDLIEQITDAENGSDLYGFEKDLDESICQVNKKYLTFISSDSSSFLIIKNDQTKVISNY 471

Qy     184 LKLN-EFPIRQPFSKWFVSKSEY 205
           :     |  |    :| ::  ||
Db     472 VWAQLYFETR----RWIINDIEY 490


Search completed: April 14, 2003, 12:12:18
Job time : 18 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 12:08:21 ; Search time 28 Seconds
(without alignments)
325.754 Million cell updates/sec

Title: US-09-914-152-1
Perfect score: 1657
Sequence: 1 MAFPKMRLMYICLLVLGALC.....TLLDYWQALENSRGEDCPPV 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	589	35.5	326	2	US-09-055-097-6	Sequence 6, Appli
2	388	23.4	378	2	US-09-055-097-1	Sequence 1, Appli
3	359.5	21.7	378	4	US-09-482-180A-2	Sequence 2, Appli
4	350	21.1	325	2	US-09-055-097-5	Sequence 5, Appli
5	314	18.9	397	4	US-09-459-133-2	Sequence 2, Appli
6	307.5	18.6	389	4	US-09-459-133-13	Sequence 13, Appl
7	138	8.3	372	1	US-08-207-904-10	Sequence 10, Appl
8	81.5	4.9	1513	5	PCT-US93-03076-2	Sequence 2, Appli
9	79.5	4.8	612	3	US-09-212-971-14	Sequence 14, Appl
10	79.5	4.8	612	4	US-08-800-929A-14	Sequence 14, Appl
11	79.5	4.8	612	4	US-08-569-749-14	Sequence 14, Appl
12	79.5	4.8	612	4	US-09-617-053A-14	Sequence 14, Appl
13	79.5	4.8	612	5	PCT-US96-12860-14	Sequence 14, Appl
14	79	4.8	740	3	US-08-276-968A-18	Sequence 18, Appl
15	78	4.7	438	5	PCT-US95-05922A-2	Sequence 2, Appli
16	78	4.7	618	4	US-08-569-749-2	Sequence 2, Appli
17	78	4.7	618	4	US-09-069-023-29	Sequence 29, Appl
18	78	4.7	618	5	PCT-US96-12860-2	Sequence 2, Appli
19	78	4.7	741	3	US-08-276-968A-20	Sequence 20, Appl
20	77	4.6	454	3	US-08-276-968A-22	Sequence 22, Appl
21	76	4.6	740	1	US-08-016-863-16	Sequence 16, Appl
22	76	4.6	740	3	US-08-276-968A-16	Sequence 16, Appl
23	76	4.6	923	4	US-09-397-885-1	Sequence 1, Appli
24	75.5	4.6	331	4	US-09-134-001C-3626	Sequence 3626, Ap
25	75.5	4.6	821	1	US-09-377-465A-2	Sequence 2, Appli
26	75.5	4.6	3177	2	US-08-477-451-4	Sequence 4, Appli
27	73.5	4.4	329	4	US-09-651-200-19	Sequence 19, Appl
28	73.5	4.4	340	4	US-09-651-200-2	Sequence 2, Appli
29	73.5	4.4	441	4	US-09-651-200-4	Sequence 4, Appli
30	73.5	4.4	534	4	US-09-651-200-6	Sequence 6, Appli
31	73.5	4.4	534	4	US-09-651-200-24	Sequence 24, Appl
32	73.5	4.4	802	4	US-09-433-043B-120	Sequence 120, App
33	73.5	4.4	2237	1	US-08-354-973-1	Sequence 1, Appli
34	73	4.4	1054	1	US-07-596-467-4	Sequence 4, Appli
35	73	4.4	1054	1	US-07-934-374-4	Sequence 4, Appli
36	73	4.4	1054	1	US-07-783-861C-2	Sequence 2, Appli
37	72.5	4.4	289	1	US-07-593-657-7	Sequence 7, Appli
38	72.5	4.4	289	4	US-08-942-012B-4	Sequence 4, Appli
39	72.5	4.4	503	4	US-09-144-367-2	Sequence 2, Appli
40	72.5	4.4	517	1	US-08-457-274A-2	Sequence 2, Appli
41	72.5	4.4	517	5	PCT-US95-05758-2	Sequence 2, Appli
42	72	4.3	363	3	US-08-586-165-7	Sequence 7, Appli
43	72	4.3	732	1	US-08-481-626-2	Sequence 2, Appli
44	72	4.3	732	4	US-08-989-299-4	Sequence 4, Appli
45	72	4.3	887	1	US-08-327-494A-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-09-055-097-6
; Sequence 6, Application US/09055097
; Patent No. 5955282
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  HUMAN OXIDIZED LDL RECEPTOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/055,097
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0490 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  326 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  2745735
US-09-055-097-6

  Query Match 35.5%;  Score 589;  DB 2;  Length 326;
```

```
  Best Local Similarity   38.6%;  Pred. No. 2.3e-62;
  Matches  127;  Conservative   52;  Mismatches  108;  Indels   42;  Gaps    7;

Qy      12 CLLVLGALCLYFSMYSLN-----------PFKEQSFVYKK--DGN-----------FL 45
           || ||  :|   ::: |:            ||   :   |    ||            ||
Db       7 CLYVLSVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFL 66

Qy      46 KLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE 105
             | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db      67 INEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVL 126

Qy     106 TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDY 165
           : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||||:|:|:|
Db     127 NQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDN 186

Qy     166 LTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSG 223
           |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:||
Db     187 LIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSA 245

Qy     224 DVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVCLF 278
           |||  :| |     : ||||:|||| :| |          |   |      : :|:| :
Db     246 DVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLCRY 297

Qy     279 RRIVACHFIKPRTLLDYWQALENSRGEDC 307
           ||::  | | |  :   |  : : :  |
Db     298 RRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 2
US-09-055-097-1
; Sequence 1, Application US/09055097
; Patent No. 5955282
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  HUMAN OXIDIZED LDL RECEPTOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/055,097
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0490 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  DUODNOT02
;      CLONE:  1705085
US-09-055-097-1

  Query Match             23.4%;  Score 388;  DB 2;  Length 378;
  Best Local Similarity   29.0%;  Pred. No. 4.9e-38;
  Matches  104;  Conservative   60;  Mismatches  129;  Indels   66;  Gaps   10;

Qy       7 RLMYICLLV-----------LGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLP------D 49
           ||:   ||:            ||   |  |: || |           |  | ||
Db       8 RLLLAALLLVIVWTLFGPSGLGEELLSLSLASLLP------APASPGPPLALPRLLIPNQ 61

Qy      50 TDCR--QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSS----- 102
             |     ||||::||  :: : | :| ||| :||  |  :| :::| ||||  ::
Db      62 EACSGPGAPPFLLILVCTAPENLNQRNAIRASWGGLREARGLRVQTLFLLGEPNAQHPVW 121

Qy     103 AAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFIN 162
           ::  ::  ||   |||:|  | | | |||||: |: |  : || | :|:||| |:::|
Db     122 GSQGSDLASESAAQGDILQAAFQDSYRNLTLKTLSGLNWAEKHCPMARYVLKTDDDVYVN 181

Qy     163 V-DYLTELLLKKNRTTRF-------------------------FTGFLKLNEFPIRQPF 195
           | : ::||:|:  |  ::                         : | :     | | |
Db     182 VPELVSELVLRGGRWGQWERSTEPQREAEQEGGQVLHSEEVPLLYLGRVHWRVNPSRTPG 241

Qy     196 SKWFVSKSEYP--WDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERL 253
           :  ||: ::|  |  :||: |||||| |      :  |:   | : |||||||:   |
Db     242 GRHRVSEEQWPHTWGPFPPYASGTGYVLSASAVQLILKVASRAPLLPLEDVFVGVSARRG 301

Qy     254 NIRLEELHSQPTFFPGGLRFSV---CLFRRIVACHFIKPRTLLDYWQALENSRGEDCPP 309
           :      :|    |   : :  |  : ::  | : |  : : |: :  | ||   |
Db     302 GL----APTQCVKLAGATHYPLDRCCYGKFLLTSHRLDPWKMQEAWKLVGGSDGERTAP 356


RESULT 3
US-09-482-180A-2
; Sequence 2, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Gao, Zeren
;  APPLICANT:  Whitmore, Theodore E.
;  APPLICANT:  Jaspers, Stephen
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
;  TITLE OF INVENTION:  BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZNSSP6
;  FILE REFERENCE: 98-80
;  CURRENT APPLICATION NUMBER: US/09/482,180A
;  CURRENT FILING DATE:  2000-01-12
;  PRIOR APPLICATION NUMBER: 60/115,721
;  PRIOR FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 378
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-482-180A-2

  Query Match             21.7%;  Score 359.5;  DB 4;  Length 378;
  Best Local Similarity   32.3%;  Pred. No. 1.3e-34;
  Matches   92;  Conservative   45;  Mismatches  137;  Indels   11;  Gaps    6;

Qy      30 PFKEQSFV-YKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGK-ERMVK 87
           | : : |: |:   ||  | :        ||:| : |    :  | ||| |||:    :
Db      90 PSRHRLFLTYRHCRNFSILLEPSGCSKDTFLLLAIKSQPGHVERRAAIRSTWGRVGGWAR 149

Qy      88 GKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 147
           |:|||  ||||   ||   : :  ||:   ||:| || : ::||||| :    ||   ||
Db     150 GRQLKLVFLLGVAGSAPPAQLLAYESREFDDILQWDFTEDFFNLTLKELHLQRWVVAACP 209

Qy     148 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 207
           || |::| | |:|::|  : | |   :        | :     | |    |:|:  | |
Db     210 QAHFMLKGDDDVFVHVPNVLEFLDGWDPAQDLLVGDVIRQALPNRNTKVKYFIPPSMYRA 269
```

```
Qy    208 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 267
          |||:  | ||| |     ::  : :    : ::||||||:|| ||   |  :|    |
Db    270 THYPPYAGGGGYVMSRATVRRLQAIMEDAELLSIDDVFVGMCLRRLG--LSPMH-HAGFK 326

Qy    268 PGGLR-----FSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDC 307
          |:|        ||:| ::  | : |  :   | ||    |  |
Db    327 TFGIRRPLDPLDPCLYRGLLLVHRLSPLEMWTMW-ALVTDEGLKC 370


RESULT 4
US-09-055-097-5
; Sequence 5, Application US/09055097
; Patent No. 5955282
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  HUMAN OXIDIZED LDL RECEPTOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/055,097
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0490 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1150971
US-09-055-097-5

  Query Match             21.1%;  Score 350;  DB 2;  Length 325;
  Best Local Similarity   29.0%;  Pred. No. 1.5e-33;
  Matches   92;  Conservative   56;  Mismatches  135;  Indels   34;  Gaps   10;

Qy      7 RLMYICLLVLGAL-----C-LYFSMYSLNPFKEQSFVYKKDGN-------------FLK 46
            :|:  |||||  :     | |   :: ||   |: | |  : :             :|:
Db      7 KLLLRCLLVLPLILLVDYCGLLTHLHELN--FERHFHYPLNDDTGSGSASSGLDKFAYLR 64

Qy     47 LPDTDCR---QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 103
          :|         | | :|: |:      | |||:||| |       |:  |||||   :
Db     65 VPSFTAEVPVDQPARLTMLIKSAVGNSRRREAIRRTWGYEGRFSDVHLRRVFLLGTAEDS 124

Qy    104 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 163
            |:|  ||: ||||:| || | |:| |||||:|: |      :: | :  | | :::
```

```
Db    125 --EKDVAWESREHGDILQADFTDAYFNNTLKTMLGMRWASEQFNRSEFYLFVDDDYYVSA 182

Qy    164 DYLTELL---LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYV 220
            : : |    : ::    | | :     |:|  ||||:||  |||:||:||: :   ::
Db    183 KNVLKFLGRGRQSHQPELLFAGHV-FQTSPLRHKFSKWYVSLEEYPFDRWPPYVTAGAFI 241

Qy    221 FSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRR 280
          |     |:|  |  :|  : :||::|:   :  | |:  |     |          :
Db    242 LSQKALRQLYAASVHLPLFRFDDVYLGIVALKAGISLQ--HCDDFRFHRPAYKGPDSYSS 299

Qy    281 IVACH-FIKPRTLLDYW 296
          ::| | |  |  :   |
Db    300 VIASHEFGDPEEMTRVW 316


RESULT 5
US-09-459-133-2
; Sequence 2, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Jaspers, Stephen R.
;  APPLICANT:  Gao, Zeren
;  TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
;  FILE REFERENCE: 98-77
;  CURRENT APPLICATION NUMBER: US/09/459,133
;  CURRENT FILING DATE:  1999-12-10
;  PRIOR APPLICATION NUMBER: 60/111,697
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 397
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (137)...(137)
;   OTHER INFORMATION: Xaa is Gly or Ser
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(397)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-459-133-2

  Query Match             18.9%;  Score 314;  DB 4;  Length 397;
  Best Local Similarity   29.9%;  Pred. No. 4.5e-29;
  Matches   78;  Conservative   50;  Mismatches  117;  Indels   16;  Gaps    6;

Qy     49 DTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKE 108
          |||     |:|:| | |   : ||| |:|:|||          ::  ||||:    |   :
Db    144 DTDV----PYLLLAVKSEPGRFAERQAVRETWGS----PAPGIRLLFLLGSPVGEA-GPD 194

Qy    109 VDQ----ESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVD 164
          :|     ||:|: |::  ||||| :| ||| :: : |: | ||  :||::   | |::
Db    195 LDSLVAWESRRYSDLLLWDFLDVPFNQTLKDLLLLAWLGRHCPTVSFVLRAQDDAFVHTP 254

Qy    165 YLTELL--LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFS 222
          |   |  |        : | :     |:|:|   ::| :| :    || : || ||| :
Db    255 ALLAHLRALPPASARSLYLGEVFTQAMPLRKPGGPFYVPESFFE-GGYPAYASGGGYVIA 313

Qy    223 GDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIV 282
          | :|  :   :  |     |||: |||:  | :  :       | :|       | || ::
Db    314 GRLAPWLLRAAARVAPFPFEDVYTGLCIRALGLVPQAHPGFLTAWPADRTADHCAFRNLL 373

Qy    283 ACHFIKPRTLLDYWQALENSR 303
          : |:  :  |: |:: |
Db    374 LVRPLGPQASIRLWKQLQDPR 394


RESULT 6
US-09-459-133-13
```

```
; Sequence 13, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Jaspers, Stephen R.
;  APPLICANT:  Gao, Zeren
;  TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
;  FILE REFERENCE: 98-77
;  CURRENT APPLICATION NUMBER: US/09/459,133
;  CURRENT FILING DATE:  1999-12-10
;  PRIOR APPLICATION NUMBER: 60/111,697
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
;   LENGTH: 389
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-459-133-13

  Query Match             18.6%;  Score 307.5;  DB 4;  Length 389;
  Best Local Similarity   29.6%;  Pred. No. 2.6e-28;
  Matches   86;  Conservative   45;  Mismatches  135;  Indels   25;  Gaps    8;

Qy      15 VLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCR-QTPPFLVLLVTSSHKQLAER 73
           :| | |  | ::   |  | |             |   |  :  |:|:| | |     | |
Db     111 LLSAACRSFPLWL--PAGEGS-----------PVASCSDKDVPYLLLAVKSEPGHFAAR 156

Qy      74 MAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYN 130
            |:|:|||    | | :|   ||||:             |  ||:|:||::  ||||| ||
Db     157 QAVRETWGSP--VAGTRL--LFLLGSPLGMGGPDLRSLVTWESRRYGDLLLWDFLDVPYN 212

Qy     131 LTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL--LKKNRTTRFFTGFLKLNE 188
           ||| :: : |:   ||   ||::   | |:::  | | |  |        : | :
Db     213 RTLKDLLLLTWLSHHCPDVNFVLQVQDDAFVHIPALLEHLQTLPPTWARSLYLGEIFTQA 272

Qy     189 FPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGL 248
           |:|:|   ::| |: :  | || : || ||| || :|  :   :  |     :||: |
Db     273 KPLRKPGGPFYVPKTFFEGD-YPAYASGGGYVISGRLAPWLLQAAARVAPFPFDDVYTGF 331

Qy     249 CLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 299
           |   | :         | :|       |  | ::  | : |:  :  |: |
Db     332 CFRALGLAPRAHPGFLTAWPAERTRDPCAVRGLLLVHPVSPQDTIWLWRHL 382


RESULT 7
US-08-207-904-10
; Sequence 10, Application US/08207904
; Patent No. 5477002
;  GENERAL INFORMATION:
;    APPLICANT:  Tuttle, AnnMarie
;    APPLICANT:  Crossland, Lyle D.
;    TITLE OF INVENTION:  Anther-Specific cDNA Sequences, Genomic
;    TITLE OF INVENTION:  DNA Sequences and Recombinant DNA Sequences
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/207,904
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/908,242
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lazar, Steven R.
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  CGC 1624
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8615
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  372 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-207-904-10

  Query Match             8.3%;  Score 138;  DB 1;  Length 372;
  Best Local Similarity   23.7%;  Pred. No. 6.6e-08;
  Matches   53;  Conservative   41;  Mismatches   86;  Indels   44;  Gaps    9;

Qy      58 FLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTF---------FLLG--TTSSAAET 106
           |:|: : ::      | ::|:||    | || :|:           |::|   |
Db     103 FIVVGINTAFSSRKRRDSLRETW----MPKGDKLRKLEKEKGIVIRFVIGHSATRGGVLD 158

Qy     107 KEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIE-----WVHRFCPQAAFVMKTDSDMFI 161
           : :| |  :: | :: | :: |: |: || :        |       | | :| | |: :
Db     159 RAIDSEDAQYKDFLRLDHVEGYHELSTKTRLYFSKAVSIW------DADFYVKVDDDVHL 212

Qy     162 NVDYLTELLLKKNRTTRFFTGFLK----LNEFPIRQPFSKWFVSKSEYPW------DRYP 211
           |:  |   | |     | : | :|    |::  :|         : || |      ::|
Db     213 NLGMLANTLAKYKSKPRVYIGCMKSGPVLSQKGVR-------YYEPEY-WKFGEEGNKYF 264

Qy     212 PFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNI 255
           :|  |  | |:|| :   |  :     ||| :|  |  | :
Db     265 RHATGQIYGISRDLASYISINSGILHRYANEDVSLGSWLIGLEV 308


RESULT 8
PCT-US93-03076-2
; Sequence 2, Application PC/TUS9303076
;  GENERAL INFORMATION:
;    APPLICANT:  Whitehead Institute for Biomedical Research
;    TITLE OF INVENTION:  GAP-Associated Protein p190 and
;    TITLE OF INVENTION:  Transduction
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03076
;      FILING DATE:  19930331
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI92-03A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1513 amino acids
;      TYPE:  AMINO ACID
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-03076-2

  Query Match             4.9%;  Score 81.5;  DB 5;  Length 1513;
  Best Local Similarity   23.6%;  Pred. No. 3.9;
  Matches   37;  Conservative   22;  Mismatches   63;  Indels   35;  Gaps    5;

Qy      72 ERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNL 131
           ::: : | : :::|  || |   ||||   |    :  |          |  |:   ||
Db     137 DQLGLEQDFEQKQMPDGKLLVDGFLLGIDVSRGMNRNFDD---------QLKFVSNLYNQ 187

Qy     132 TLKTMMGI--------EWVHRFCPQA---------AFVMKTDSDMFINVDY----LTELL 170
             ||   |        | | |:   |           |::| :   :|||     | :|:
Db     188 LAKTKKPIVIVLTKCDEGVERYIRDAHTFALSKKNLQVVETSARSNVNVDLAFTTLVQLI 247

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 207
           |    |:    |  |     :|   :   :| :| |
Db     248 DKSRGKTKIIPYFEAL-----KQQSQQIATAKDKYEW 279


RESULT 9
US-09-212-971-14
; Sequence 14, Application US/09212971B
; Patent No. 6107041
; GENERAL INFORMATION:
;  APPLICANT: Korneluk, Robert G
;  APPLICANT: MacKenzie, Alexander E
;  APPLICANT: Liston, Peter
;  APPLICANT: Baird, Stephen
;  APPLICANT: Tsang, Benjamin K
;  APPLICANT: Pratt, Christine
;  TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
;  TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
;  TITLE OF INVENTION: DISEASE
;  FILE REFERENCE: 07891/009002
;  CURRENT APPLICATION NUMBER: US/09/212,971B
;  CURRENT FILING DATE: 1998-12-16
;  EARLIER APPLICATION NUMBER: 60/017,354
;  EARLIER FILING DATE: 1996-04-26
;  EARLIER APPLICATION NUMBER: 60/030,590
;  EARLIER FILING DATE: 1996-11-14
;  EARLIER APPLICATION NUMBER: 08/800,929
;  EARLIER FILING DATE: 1997-02-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 612
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-212-971-14

  Query Match             4.8%;  Score 79.5;  DB 3;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.6;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      59 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 108
           :|:: |   |   |    |  :|||  :: :  |:  :|        |         | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy     109 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 161
             |    ||  :|:|: : ::  ||      |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     162 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 213
                   : :  :|:| |      |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK--------------------------- 523

Qy     214 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270
                 :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     271 LRFSVCLFRRI 281
```

```
           ||: | :
Db     565 --CKVCMDREV 573


RESULT 10
US-08-800-929A-14
; Sequence 14, Application US/08800929A
; Patent No. 6133437
;  GENERAL INFORMATION:
;    APPLICANT:  Korneluk, Robert G
;    APPLICANT:  MacKenzie, Alexander E
;    APPLICANT:  Liston, Peter
;    APPLICANT:  Baird, Stephen
;    APPLICANT:  Tsang, Benjamin K
;    APPLICANT:  Pratt, Christine
;    TITLE OF INVENTION:  DETECTION AND MODULATION OF
;    TITLE OF INVENTION:  IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERA
;    TITLE OF INVENTION:  DISEASE
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/800,929A
;      FILING DATE:  13-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/030,590
;      FILING DATE:  14-NOV-1996
;      APPLICATION NUMBER:  60/017,354
;      FILING DATE:  26-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:  07891/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-428-0200
;      TELEFAX:  617-428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-800-929A-14

  Query Match             4.8%;  Score 79.5;  DB 4;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.6;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      59 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 108
           :|:: |   |   |    |  :|||  :: :  |:  :|        |         | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy     109 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 161
             |    ||  :|:|: : ::  ||      |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     162 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 213
                   : :  :|:| |      |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK--------------------------- 523
```

```
Qy     214 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270
                 :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     271 LRFSVCLFRRI 281
           ||: | :
Db     565 --CKVCMDREV 573


RESULT 11
US-08-569-749-14
; Sequence 14, Application US/08569749
; Patent No. 6187557
;  GENERAL INFORMATION:
;    APPLICANT:  Rothe, Mike
;    APPLICANT:  Goeddel, David V
;    TITLE OF INVENTION:  INHIBITORS OF APOPTOSIS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/569,749
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brezner, David J.
;      REGISTRATION NUMBER:  24,774
;      REFERENCE/DOCKET NUMBER:  A-62464/DJB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)781-1989
;      TELEFAX:  (415)398-3249
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-569-749-14

  Query Match             4.8%;  Score 79.5;  DB 4;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.6;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      59 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 108
           :|:: |   |   |    |  :||| :: :  |:  :|        |        | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy     109 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 161
             |    ||  :|:|: : ::  ||     |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     162 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 213
                   : :  :|:| |     |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK---------------------------- 523

Qy     214 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270
                 :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     271 LRFSVCLFRRI 281
           ||: | :
Db     565 --CKVCMDREV 573


RESULT 12
US-09-617-053A-14
; Sequence 14, Application US/09617053A
; Patent No. 6300492
; GENERAL INFORMATION:
;  APPLICANT: Korneluk, Robert G
;  APPLICANT:  MacKenzie, Alexander E
;  APPLICANT:  Liston, Peter
;  APPLICANT:  Baird, Stephen
;  APPLICANT:  Tsang, Benjamin K
;  APPLICANT:  Pratt, Christine
;  TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
;  TITLE OF INVENTION:  NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
;  TITLE OF INVENTION:  DISEASE
;  FILE REFERENCE: 07891/009003
;  CURRENT APPLICATION NUMBER: US/09/617,053A
;  CURRENT FILING DATE:  2000-07-14
;  PRIOR APPLICATION NUMBER: US 08/800,929
;  PRIOR FILING DATE: 1997-02-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 612
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-617-053A-14

  Query Match             4.8%;  Score 79.5;  DB 4;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.6;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      59 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 108
           :|:: |   |   |    |  :||| :: :  |:  :|        |        | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy     109 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 161
             |    ||  :|:|: : ::  ||     |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     162 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 213
                   : :  :|:| |     |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK---------------------------- 523

Qy     214 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270
                 :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     271 LRFSVCLFRRI 281
           ||: | :
Db     565 --CKVCMDREV 573


RESULT 13
PCT-US96-12860-14
; Sequence 14, Application PC/TUS9612860
;  GENERAL INFORMATION:
;    APPLICANT:  TULARIK, INC.
;    TITLE OF INVENTION:  INHIBITORS OF APOPTOSIS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/12860
;      FILING DATE:  06 AUG 1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. Serial Nos. 08/512,946 & 08/569,749
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brezner, David J.
;      REGISTRATION NUMBER:  24,774
;      REFERENCE/DOCKET NUMBER:  A-62464/DJB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)781-1989
;      TELEFAX:  (415)398-3249
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US96-12860-14

  Query Match             4.8%;  Score 79.5;  DB 5;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.6;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      59 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 108
           :|:: |   |   |    |  :|||  :: :  |:  :|        |         | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy     109 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 161
             |    ||  :|:|: : ::  ||     |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     162 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 213
                     : :  :|:| |     |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK--------------------------- 523

Qy     214 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270
                   :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     271 LRFSVCLFRRI 281
              ||: | :
Db     565 --CKVCMDREV 573


RESULT 14
US-08-276-968A-18
; Sequence 18, Application US/08276968A
; Patent No. 6015701
;  GENERAL INFORMATION:
;    APPLICANT:  Pierce, James M.
;    APPLICANT:  Shoreibah, Mohamed G.
;    APPLICANT:  Adler, Beverly
;    APPLICANT:  Fregien, Nevis L.
;    TITLE OF INVENTION:  N-Acetylglucosaminyltransferase V
;    TITLE OF INVENTION:  Proteins and Sequences
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,968A
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/905,795
;      FILING DATE:  29-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/016,863
;      FILING DATE:  10-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  34-92D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303)-499-8080
;      TELEFAX:  (303)-499-8089
;      TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 740 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-276-968A-18

  Query Match             4.8%;  Score 79;  DB 3;  Length 740;
  Best Local Similarity   18.4%;  Pred. No. 2.5;
  Matches   58;  Conservative   48;  Mismatches  104;  Indels  106;  Gaps   14;

Qy      25 MYSLNPFK--EQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGK 82
           :: :|  |   || || |  :|                                    | |
Db     436 IHHINEIKRQNQSLVYGKVDSF--------------------------------W-K 459

Qy      83 ERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMG---- 138
            : :   : |:  :  |  : || :   : || :  :|  :    |  :|
Db     460 NKKIYLDIIHTYMEVHATVYGSSTKNIPSYVKNHGILSGRDLQFLLRETKLFVGLGFPYE 519

Qy     139 ----IEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQP 194
               :| :  |             |:|  :      | :: | || |   | |   : |
Db     520 GPAPLEAIANGC-----------AFLNPKFSPP---KSSKNTDFFIGKPTLRELTSQHP 564

Qy     195 FSKWFVSKSEYPWDRYPPFCSGTGYVFSGDV--ASQVYNVSKSV------PYIKLEDVFV 246
           ::: |:               |  :|:: |:    :| :  |::      ||:  |
Db     565 YAEVFI--------------GRPHVWTVDLNNREEVEDAVKAILNQKIEPYMPYEFTCE 609

Qy     247 GLCLERLNIRLEE---LHSQPTF----------FPGGLRFSVCLFRRIVAC------HF 286
           |: |:|:|  :|:    | |  :           ||     ||   ::: |      |
Db     610 GM-LQRINAFIEKQDFCHGQVMWPPLSALQVKLAEPGQSCKQVCQENQLI-CEPSFFQHL 667

Qy     287 IKPRTLLDYWQALENS 302
           | : || |    ::|
Db     668 NKEKDLLKYRVTCQSS 683


RESULT 15
PCT-US95-05922A-2
; Sequence 2, Application PC/TUS9505922A
;  GENERAL INFORMATION:
;    APPLICANT:  HE, ET AL.
;    TITLE OF INVENTION:  Human Inhibitor of Apoptosis Gene 1
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
```

```
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/05922A
;      FILING DATE:  11 MAY 1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-292
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
PCT-US95-05922A-2

  Query Match             4.7%;  Score 78;  DB 5;  Length 438;
  Best Local Similarity   19.3%;  Pred. No. 1.5;
  Matches   52;  Conservative   46;  Mismatches   88;  Indels   84;  Gaps   11;

Qy      32 KEQSFVYKKDGNF-------LKLPDTDCRQT--PPFL-----------VLLVTSSHKQL 70
           | | || :  | :        |   ||   :   || :           |:: |   |
Db     160 KGQEFVDEIQGRYPHLLEQLLSTSDTTGEENADPPIIHFGPGESSSEDAVMMNTPVVKSA 219

Qy      71 AE----RMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGD-------- 118
            |    |  ::||   : :  |:  ||   : :    || :: ::| ::  :
Db     220 LEMGFNRDLVKQTVQSKILTTGENYKTVNDIVSALLNAEDEKREEEKEKQAEEMASDDLS 279

Qy     119 IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFINV-DYLTELLLK 172
           :|:|: : ::  ||     |  ::    :::   |   ::|  : : :   : :  :|:|
Db     280 LIRKNRMALFQQLTCVLPILDNLLKANVINK---QEHDIIKQKTQIPLQARELIDTILVK 336

Qy     173 KNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN- 231
            |     |  ||                                    :: | :|
Db     337 GNAAANIFKNCLK-----------------------------------EIDSTLYKN 358

Qy     232 --VSKSVPYIKLEDVFVGLCLERLNIRLEE 259
             | |:: ||  |||  || ||    ||:|
Db     359 LFVDKNMKYIPTEDV-SGLSLEEQLRRLQE 387


Search completed: April 14, 2003, 12:11:56
Job time : 31 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 12:10:34 ; Search time 15 Seconds
(without alignments)
1263.478 Million cell updates/sec

Title: US-09-914-152-1
Perfect score: 1657
Sequence: 1 MAFPKMRLMYICLLVLGALC.....TLLDYWQALENSRGEDCPPV 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	589	35.5	326	10	US-09-804-357-8	Sequence 8, Appli
2	589	35.5	326	10	US-09-804-006-8	Sequence 8, Appli
3	569	34.3	422	10	US-09-804-357-11	Sequence 11, Appl
4	569	34.3	422	10	US-09-529-063-25	Sequence 25, Appl
5	569	34.3	422	10	US-09-804-006-11	Sequence 11, Appl
6	567	34.2	422	9	US-10-112-616A-2	Sequence 2, Appli
7	558	33.7	331	10	US-09-804-357-9	Sequence 9, Appli
8	558	33.7	331	10	US-09-804-006-9	Sequence 9, Appli
9	555.5	33.5	319	10	US-09-739-451-2	Sequence 2, Appli
10	555.5	33.5	320	10	US-09-739-451-10	Sequence 10, Appl
11	555.5	33.5	331	9	US-09-992-598-209	Sequence 209, App
12	555.5	33.5	331	9	US-09-989-293A-209	Sequence 209, App
13	555.5	33.5	331	9	US-10-063-547-36	Sequence 36, Appl
14	555.5	33.5	331	9	US-09-989-735-209	Sequence 209, App
15	555.5	33.5	331	9	US-09-990-444-209	Sequence 209, App
16	555.5	33.5	331	9	US-09-989-730-209	Sequence 209, App
17	555.5	33.5	331	9	US-09-990-436-209	Sequence 209, App
18	555.5	33.5	331	9	US-09-991-181-209	Sequence 209, App
19	555.5	33.5	331	9	US-09-993-687-209	Sequence 209, App
20	555.5	33.5	331	9	US-09-989-734-209	Sequence 209, App
21	555.5	33.5	331	9	US-09-997-653-209	Sequence 209, App
22	555.5	33.5	331	9	US-10-174-590-162	Sequence 162, App
23	555.5	33.5	331	9	US-10-176-758-162	Sequence 162, App
24	555.5	33.5	331	9	US-10-063-616-36	Sequence 36, Appl
25	555.5	33.5	331	9	US-10-175-737-162	Sequence 162, App
26	555.5	33.5	331	9	US-09-993-667-209	Sequence 209, App
27	555.5	33.5	331	9	US-10-063-502-36	Sequence 36, Appl
28	555.5	33.5	331	9	US-10-173-706-162	Sequence 162, App
29	555.5	33.5	331	9	US-10-175-738-162	Sequence 162, App
30	555.5	33.5	331	9	US-10-175-752-162	Sequence 162, App
31	555.5	33.5	331	9	US-10-176-482-162	Sequence 162, App
32	555.5	33.5	331	9	US-10-176-757-162	Sequence 162, App
33	555.5	33.5	331	9	US-10-176-913-162	Sequence 162, App
34	555.5	33.5	331	9	US-10-180-552-162	Sequence 162, App
35	555.5	33.5	331	9	US-10-180-557-162	Sequence 162, App
36	555.5	33.5	331	9	US-09-990-438-209	Sequence 209, App
37	555.5	33.5	331	9	US-09-990-562-209	Sequence 209, App
38	555.5	33.5	331	9	US-09-997-428-209	Sequence 209, App
39	555.5	33.5	331	9	US-09-997-666-209	Sequence 209, App
40	555.5	33.5	331	9	US-10-173-700-162	Sequence 162, App
41	555.5	33.5	331	9	US-10-174-572-162	Sequence 162, App
42	555.5	33.5	331	9	US-10-174-579-162	Sequence 162, App
43	555.5	33.5	331	9	US-10-174-582-162	Sequence 162, App
44	555.5	33.5	331	9	US-10-174-588-162	Sequence 162, App
45	555.5	33.5	331	9	US-10-175-739-162	Sequence 162, App

ALIGNMENTS

RESULT 1
US-09-804-357-8
; Sequence 8, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-357-8

Query Match 35.5%; Score 589; DB 10; Length 326;
Best Local Similarity 38.6%; Pred. No. 9e-50;
Matches 127; Conservative 52; Mismatches 108; Indels 42; Gaps 7;

```
Qy      12 CLLVLGALCLYFSMYSLN----------PFKEQSFVYKK--DGN----------FL 45
           || ||  :|   ::: |:           ||   :   |    ||           ||
Db       7 CLYVLSVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFL 66

Qy      46 KLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE 105
             | :  ||||:|::::||:   | |||:||| |  || :: | ||||  :
Db      67 INEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVL 126

Qy     106 TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDY 165
             : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||||:|:|:|
Db     127 NQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDN 186
```

```
Qy    166 LTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSG 223
          |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:||
Db    187 LIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSA 245

Qy    224 DVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVCLF 278
          |||  :|  |     : ||||:|||| :| |          |   |     : :|:| :
Db    246 DVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLCRY 297

Qy    279 RRIVACHFIKPRTLLDYWQALENSRGEDC 307
          ||::  | | |  :   |  : : :   |
Db    298 RRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 2
US-09-804-006-8
; Sequence 8, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
;  APPLICANT: White, David
;  APPLICANT:  Zhou, Jianghong
;  APPLICANT:  Tartaglia, Louis A.
;  TITLE OF INVENTION: LEPTIN INDUCED GENES
;  FILE REFERENCE: 07334/126001
;  CURRENT APPLICATION NUMBER: US/09/804,006
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/292,228
;  PRIOR FILING DATE: 1999-04-15
;  PRIOR APPLICATION NUMBER: US 60/108,379
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 09/150,857
;  PRIOR FILING DATE: 1998-09-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 326
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-804-006-8

  Query Match             35.5%;  Score 589;  DB 10;  Length 326;
  Best Local Similarity   38.6%;  Pred. No. 9e-50;
  Matches  127;  Conservative   52;  Mismatches  108;  Indels   42;  Gaps    7;

Qy     12 CLLVLGALCLYFSMYSLN-----------PFKEQSFVYKK--DGN-----------FL 45
          || ||  :|   ::: |:            ||   :   |    ||            ||
Db      7 CLYVLSVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFL 66

Qy     46 KLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE 105
             | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db     67 INEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVL 126

Qy    106 TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDY 165
           : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||||:|:|:|
Db    127 NQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDN 186

Qy    166 LTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSG 223
          |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:||
Db    187 LIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSA 245

Qy    224 DVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVCLF 278
          |||  :|  |     : ||||:|||| :| |          |   |     : :|:| :
Db    246 DVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLCRY 297

Qy    279 RRIVACHFIKPRTLLDYWQALENSRGEDC 307
          ||::  | | |  :   |  : : :   |
Db    298 RRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 3
US-09-804-357-11
; Sequence 11, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
;  APPLICANT: White, David
;  APPLICANT:  Zhou, Jianghong
;  APPLICANT:  Tartaglia, Louis A.
;  TITLE OF INVENTION: LEPTIN INDUCED GENES
;  FILE REFERENCE: 07334/109001
;  CURRENT APPLICATION NUMBER: US/09/804,357
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/195,896
;  PRIOR FILING DATE: 1998-11-19
;  PRIOR APPLICATION NUMBER: US 60/108,379
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 09/150,857
;  PRIOR FILING DATE: 1998-09-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-804-357-11

  Query Match             34.3%;  Score 569;  DB 10;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 1.2e-47;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy     52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
          |::  |||:||: :    |:  | |||||| | :  | |:   ||| :   :    : :
Db    146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy    110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
           :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db    206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy    170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
          ||| :   |  :|||:|    | |   |||::    || :||| |||||||||||:|
Db    266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy    228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
          ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  :: |
Db    326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy    288 KPRTLLDYWQALENSRGEDC 307
          :|  |: ||  |: ::   |
Db    386 QPSELIKYWNHLQQNKHNAC 405


RESULT 4
US-09-529-063-25
; Sequence 25, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
;  APPLICANT: FUKUSHIMA, DAIKICHI
;  APPLICANT:  SHIBAYAMA, SHIRO
;  APPLICANT:  TADA, HIDEAKI
;  TITLE OF INVENTION: POLYPEPTIDE, cDNA ENCODING THE POLYPEPTIDE, AND USE OF
;  TITLE OF INVENTION:  THE BOTH
;  FILE REFERENCE: Q58769
;  CURRENT APPLICATION NUMBER: US/09/529,063
;  CURRENT FILING DATE:  2000-04-07
;  PRIOR APPLICATION NUMBER: PCT/JP98/04514
;  PRIOR FILING DATE: 1998-10-06
;  PRIOR APPLICATION NUMBER: JP 9-274674
;  PRIOR FILING DATE: 1997-10-07
;  NUMBER OF SEQ ID NOS: 117
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-529-063-25

  Query Match             34.3%;  Score 569;  DB 10;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 1.2e-47;
```

Matches 108; Conservative 53; Mismatches 95; Indels 4; Gaps 2;

```
Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |:: .|||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
            :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405
```

RESULT 5
US-09-804-006-11
; Sequence 11, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-804-006-11

Query Match 34.3%; Score 569; DB 10; Length 422;
Best Local Similarity 41.5%; Pred. No. 1.2e-47;
Matches 108; Conservative 53; Mismatches 95; Indels 4; Gaps 2;

```
Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
            :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405
```

RESULT 6
US-10-112-616A-2
; Sequence 2, Application US/10112616A
; Publication No. US20030005477A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING BETA 3 GalT2
; TITLE OF INVENTION: GENE DISRUPTIONS
; FILE REFERENCE: R-031
; CURRENT APPLICATION NUMBER: US/10/112,616A
; CURRENT FILING DATE: 2002-09-05
; PRIOR APPLICATION NUMBER: US 60/280,362
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/326,700
; PRIOR FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-112-616A-2

Query Match 34.2%; Score 567; DB 9; Length 422;
Best Local Similarity 41.2%; Pred. No. 1.8e-47;
Matches 107; Conservative 53; Mismatches 96; Indels 4; Gaps 2;

```
Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 205

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
            :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     206 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 265

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405
```

RESULT 7
US-09-804-357-9
; Sequence 9, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Mus musculus

US-09-804-357-9

Query Match 33.7%; Score 558; DB 10; Length 331;
Best Local Similarity 36.7%; Pred. No. 1e-46;
Matches 114; Conservative 58; Mismatches 127; Indels 12; Gaps 3;

```
Qy       1 MAFPKMRLMYICLLVLGALCL-YFSMYSLNP--------FKEQSFVYKKDGNFLKLPDTD 51
Db      13 MSLRSLKWSLLLLSLLSFLVIWYLSLPHYNVIERVNWMYFYEYEPIYRQDFRFTLREHSN 72

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKE 108
Db      73 CSHQNPFLVILVTSRPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQQAEREDKTLALS 132

Qy     109 VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTE 168
Db     133 LEDEHVLYGDIIRQDFLDTYNNLTLKTIMAFRWVMEFCPNAKYIMKTDTDVFINTGNLVK 192

Qy     169 LLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQ 228
Db     193 YLLNLNHSEKFFTGYPLIDNYSYRGFFHKNHISYQEYPFKVFPPYCSGLGYIMSGDLVPR 252

Qy     229 VYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIK 288
Db     253 VYEMMSHVKPIKFEDVYVGICLNLLKVDIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFS 312

Qy     289 PRTLLDYWQAL 299
Db     313 SKEIITFWQVM 323
```

RESULT 8
US-09-804-006-9
; Sequence 9, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/126001
; CURRENT APPLICATION NUMBER: US/09/804,006
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/292,228
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-006-9

Query Match 33.7%; Score 558; DB 10; Length 331;
Best Local Similarity 36.7%; Pred. No. 1e-46;
Matches 114; Conservative 58; Mismatches 127; Indels 12; Gaps 3;

```
Qy       1 MAFPKMRLMYICLLVLGALCL-YFSMYSLNP--------FKEQSFVYKKDGNFLKLPDTD 51
Db      13 MSLRSLKWSLLLLSLLSFLVIWYLSLPHYNVIERVNWMYFYEYEPIYRQDFRFTLREHSN 72

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKE 108
Db      73 CSHQNPFLVILVTSRPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQQAEREDKTLALS 132

Qy     109 VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTE 168
Db     133 LEDEHVLYGDIIRQDFLDTYNNLTLKTIMAFRWVMEFCPNAKYIMKTDTDVFINTGNLVK 192

Qy     169 LLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQ 228
Db     193 YLLNLNHSEKFFTGYPLIDNYSYRGFFHKNHISYQEYPFKVFPPYCSGLGYIMSGDLVPR 252

Qy     229 VYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIK 288
Db     253 VYEMMSHVKPIKFEDVYVGICLNLLKVDIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFS 312

Qy     289 PRTLLDYWQAL 299
Db     313 SKEIITFWQVM 323
```

RESULT 9
US-09-739-451-2
; Sequence 2, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3
; FILE REFERENCE: PF464
; CURRENT APPLICATION NUMBER: US/09/739,451
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/213,364
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/108,928
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-451-2

Query Match 33.5%; Score 555.5; DB 10; Length 319;
Best Local Similarity 36.2%; Pred. No. 1.7e-46;
Matches 112; Conservative 57; Mismatches 125; Indels 15; Gaps 2;

```
Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
Db       3 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 62

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
Db      63 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 122

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
Db     123 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 182

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
Db     183 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 242

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
Db     243 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 302

Qy     291 TLLDYWQAL 299
Db     303 EIITFWQVM 311
```

RESULT 10
US-09-739-451-10
; Sequence 10, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3

```
;  FILE REFERENCE: PF464
;  CURRENT APPLICATION NUMBER: US/09/739,451
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: 09/213,364
;  PRIOR FILING DATE: 1998-12-17
;  PRIOR APPLICATION NUMBER: 60/077,687
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/108,928
;  PRIOR FILING DATE: 1998-11-17
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 320
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-739-451-10

  Query Match             33.5%;  Score 555.5;  DB 10;  Length 320;
  Best Local Similarity   36.2%;  Pred. No. 1.7e-46;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |            | |   :|::| :|     ::|
Db       3 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 62

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      63 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 122

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     123 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 182

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     183 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 242

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
           :   |  || |||:||:||  | : :         ||   :   ||  ||::| |    :
Db     243 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 302

Qy     291 TLLDYWQAL 299
             :: :|| :
Db     303 EIITFWQVM 311


RESULT 11
US-09-992-598-209
; Sequence 209, Application US/09992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C20
;  CURRENT APPLICATION NUMBER: US/09/992,598
;  CURRENT FILING DATE:  2001-11-14
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088810
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
```

; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 33.5%; Score 555.5; DB 9; Length 331;
Best Local Similarity 36.2%; Pred. No. 1.8e-46;
Matches 112; Conservative 57; Mismatches 125; Indels 15; Gaps 2;

```
Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |                | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
            |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: : |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
             :   |  || |||:||:||  | : :        ||   :   ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
             :: :|| :
Db     315 EIITFWQVM 323
```

RESULT 12
US-09-989-293A-209
; Sequence 209, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi,Avi J.
; APPLICANT: Baker,Kevin P.
; APPLICANT: Botstein,David
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Fong,Sherman
; APPLICANT: Gerber,Hanspeter
; APPLICANT: Gerritsen,Mary E.

; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,J.Christopher
; APPLICANT: Gurney,Austin L.
; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C66
; CURRENT APPLICATION NUMBER: US/09/989,293A
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24

```
;  PRIOR APPLICATION NUMBER: 60/090535
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090540
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090542
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090557
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090676
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090678
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090690
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090694
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090695
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090696
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090862
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/090863
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/091360
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091478
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091544
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091519
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091626
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091633
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091978
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/091982
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/092182
;  PRIOR FILING DATE: 1998-07-09

  Query Match             33.5%;  Score 555.5;  DB 9;  Length 331;
  Best Local Similarity   36.2%;  Pred. No. 1.8e-46;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |             | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :|||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
             :   |  || |||:||:||  | : :         ||   :   ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
           :: :|| :
Db     315 EIITFWQVM 323


RESULT 13
```

```
US-10-063-547-36
; Sequence 36, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,547
;  CURRENT FILING DATE:  2002-05-02
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 36
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-547-36

  Query Match             33.5%;  Score 555.5;  DB 9;  Length 331;
  Best Local Similarity   36.2%;  Pred. No. 1.8e-46;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |             | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :|||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
             :   |  || |||:||:||  | : :         ||   :   ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
           :: :|| :
Db     315 EIITFWQVM 323


RESULT 14
US-09-989-735-209
; Sequence 209, Application US/09989735
; Publication No. US20020193299A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
```

; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C61
; CURRENT APPLICATION NUMBER: US/09/989,735
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24

```
;  PRIOR APPLICATION NUMBER: 60/090542
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090557
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090676
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090678
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090690
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090694
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090695
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090696
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090862
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/090863
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/091360
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091478
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091544
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091519
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091626
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091633
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091978
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/091982
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/092182
;  PRIOR FILING DATE: 1998-07-09

  Query Match             33.5%;  Score 555.5;  DB 9;  Length 331;
  Best Local Similarity   36.2%;  Pred. No. 1.8e-46;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |            | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
           ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
           :  |  || |||:||:||  | : :        ||   :   ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
           :: :|| :
Db     315 EIITFWQVM 323


RESULT 15
US-09-990-444-209
; Sequence 209, Application US/09990444
; Publication No. US20020193300A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C19
;  CURRENT APPLICATION NUMBER: US/09/990,444
;  CURRENT FILING DATE:  2001-11-14
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
```

; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 33.5%; Score 555.5; DB 9; Length 331;
Best Local Similarity 36.2%; Pred. No. 1.8e-46;
Matches 112; Conservative 57; Mismatches 125; Indels 15; Gaps 2;

```
Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |                | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
```

```
          :   |  || |||:||:||  | : :        ||   :   ||  ||::| |    :
Db    255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy    291 TLLDYWQAL 299
          :: :|| :
Db    315 EIITFWQVM 323
```

Search completed: April 14, 2003, 12:18:15
Job time : 17 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:37:51 ; Search time 38 Seconds
(without alignments)
981.847 Million cell updates/sec

Title: US-09-914-152-1_COPY_31_310
Perfect score: 1498
Sequence: 1 FKEQSFVYKKDGNFLKLPDT.....TLLDYWQALENSRGEDCPPV 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1498	100.0	310	21	AAB26749	Human beta-1,3 gal
2	1492	99.6	310	21	AAY94641	Human beta3Gal-T5
3	1016	67.8	192	23	AAE15931	Human 5174397/63-2
4	577	38.5	326	15	AAR57433	Beta-1,3-galactosy
5	569	38.0	422	20	AAY02366	Polypeptide identi
6	548.5	36.6	319	20	AAY06461	Human Dendriac. H
7	548.5	36.6	331	21	AAY66685	Membrane-bound pro
8	548.5	36.6	331	22	AAU29104	Human PRO polypept
9	548.5	36.6	331	22	AAB87543	Human PRO1074. Ho
10	548.5	36.6	331	22	AAB65208	Human PRO1074 (UNQ
11	548.5	36.6	331	23	AAU80224	Human beta1,3-acet
12	543.5	36.3	331	22	AAB93992	Human protein sequ
13	497.5	33.2	193	23	AAE15927	Human 4566767/84-3
14	497.5	33.2	193	23	AAE15928	Mouse 2745735/84-3
15	480	32.0	196	23	AAE15929	Human 4502339_JBC_
16	479	32.0	196	23	AAE15930	Mouse 2745737/157-
17	443.5	29.6	195	23	AAE15933	Mouse 2745739/84-3
18	433.5	28.9	195	23	AAE15932	Human 4502343_JBC_
19	413	27.6	372	19	AAW64558	Human epidermoid c
20	413	27.6	372	22	AAU29167	Human PRO polypept
21	413	27.6	372	22	AAB88404	Human membrane or
22	413	27.6	372	22	AAB49749	Human beta 1,3-N-a
23	413	27.6	372	22	AAB49750	Human beta 1,3-N-a
24	413	27.6	372	23	ABB09716	Amino acid sequenc
25	413	27.6	372	23	AAU11272	Human beta1,3-N-ac
26	401.5	26.8	368	23	AAU11273	Murine beta1,3-N-a
27	396	26.4	378	22	AAE04752	Human beta-1,3-gal
28	396	26.4	378	23	AAE21679	Human galactosyltr
29	396	26.4	378	23	ABB06980	Human beta 1,3-N-a
30	396	26.4	378	23	ABB07522	Human drug metabol
31	393	26.2	371	20	AAW95171	Protein exhibiting
32	387.5	25.9	378	20	AAY34171	Human galactosyltr
33	387.5	25.9	378	22	AAM93908	Human polypeptide,
34	384	25.6	401	23	AAE22148	Human TRNFR-10 pro
35	381.5	25.5	574	22	ABB63976	Drosophila melanog
36	365.5	24.4	334	22	ABG20251	Novel human diagno
37	360	24.0	373	22	AAB93536	Human protein sequ
38	360	24.0	397	19	AAW80212	Human cardiac and
39	360	24.0	397	21	AAY84683	Amino acid sequenc
40	360	24.0	397	21	AAY69698	Human Brainiac pro
41	360	24.0	397	22	AAG66118	Human LIG46 polype
42	360	24.0	413	22	AAB49748	Human beta 1,3-N-a
43	359.5	24.0	402	21	AAB24035	Human PRO4397 prot
44	359.5	24.0	402	23	AAE22142	Human TRNFR-4 prot
45	359.5	24.0	402	23	ABG34046	Human Pro peptide

ALIGNMENTS

RESULT 1
AAB26749
ID AAB26749 standard; Protein; 310 AA.
XX
AC AAB26749;
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta-1,3 galactose transferase protein sequence.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system.
XX
OS Homo sapiens.
XX
PN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI; 2000-549409/50.
DR N-PSDB; AAA93875.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer -

```
XX
PS   Claim 2; Page 97-98; 123pp; Japanese.
XX
CC   This invention relates to a polypeptide (I) with beta-1,3 galactose
CC   transferase activity, or variants of (I) comprising amino acid additions,
CC   deletions and/or substitutions. Included in the invention is DNA encoding
CC   all or part of (I); expression vectors containing the DNA, host cells
CC   transformed by the vectors; a method for the preparation of the
CC   polypeptide by culture of the transformants or by expression in the milk
CC   of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC   galactose transferase protein transfers galactose by beta-1,3 bonding to
CC   N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC   GlcNAcbeta1-3Galbeta1-4Glc) to give Galbeta1-3GlcNAc. The protein and
CC   DNA encoding it are useful for the treatment and diagnosis of cancer of
CC   the digestive system. The present sequence represents the Beta-1,3
CC   galactose transferase protein.
XX
SQ   Sequence   310 AA;

  Query Match             100.0%;  Score 1498;  DB 21;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-165;
  Matches  280;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 90

Qy      61 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 150

Qy     121 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 210

Qy     181 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270

Qy     241 LRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV 280
          ||||||||||||||||||||||||||||||||||||||||
Db     271 LRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV 310


RESULT 2
AAY94641
ID   AAY94641 standard; Protein; 310 AA.
XX
AC   AAY94641;
XX
DT   15-AUG-2000  (first entry)
XX
DE   Human beta3Gal-T5 protein sequence.
XX
KW   UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;
KW   Beta3Gal-T5; human; chromosome 21q22.3; galactosylation; glycoprotein;
KW   beta1,3-galactosyl glycosylated saccharide production; glycopeptide.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          13..26
FT                   /note= "Transmembrane domain"
XX
PN   WO200029558-A1.
XX
PD   25-MAY-2000.
XX
PF   11-NOV-1999;   99WO-US26807.
XX
PR   13-NOV-1998;   98DK-0001483.
XX
PA   (CLAU/) CLAUSEN H.
XX
PI   Clausen H,  Amado M;
XX
DR   WPI; 2000-399728/34.
DR   N-PSDB; AAA27959.
XX
PT   Novel nucleic acid sequence encoding human
PT   UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
PT   useful for obtaining beta 1,3-galactosyl glycosylated saccharides and
PT   glycopeptides or glycoproteins  -
XX
PS   Claim 23; Fig 1; 74pp; English.
XX
CC   The present invention relates to a nucleic acid sequence encoding
CC   UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
CC   (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at
CC   carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence
CC   represents the human beta3Gal-T5 protein sequence. The beta3Gal-T5 gene
CC   is located on human chromosome 21q22.3. Beta3Gal-T5 is a type II
CC   transmembrane glycoprotein. The invention also relates to the beta3Gal-T5
CC   protein sequence, a nucleic acid vector comprising the beta3Gal-T5
CC   nucleotide sequence, a host cell comprising the vector, and a method for
CC   the production of the beta3Gal-T5 protein from the host cells. The
CC   methods of the invention can be used for recombinant production of
CC   beta3Gal-T5 for use as a catalyst and for recombinant production of
CC   peptides or proteins with appropriate galactosylation. The beta3Gal-T5
CC   protein can be used to obtain beta1,3-galactosyl glycosylated
CC   saccharides, glycopeptides or glycoproteins.
XX
SQ   Sequence   310 AA;

  Query Match             99.6%;  Score 1492;  DB 21;  Length 310;
  Best Local Similarity   99.6%;  Pred. No. 2.6e-164;
  Matches  279;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      31 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQ 90

Qy      61 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 150

Qy     121 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 210

Qy     181 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270

Qy     241 LRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV 280
          ||||||||||||||||||||||||||||||||||||||||
Db     271 LRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV 310


RESULT 3
AAE15931
ID   AAE15931 standard; Protein; 192 AA.
XX
AC   AAE15931;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human 5174397/63-284 protein, member of Fringe and Brainiac family.
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; human.
XX
OS   Homo sapiens.
XX
```

PN WO200187321-A2.
XX
PD 22-NOV-2001.
XX
PF 21-MAY-2001; 2001WO-IB01033.
XX
PR 19-MAY-2000; 2000GB-0012216.
XX
PA (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI Cohen S, Brueckner K, Clausen H, Keck B;
XX
DR WPI; 2002-082940/11.
XX
PT Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT treatment of disease caused by epidermal growth factor-like module
PT containing protein, such as T cell leukemia, breast cancer, stroke and
PT dementia -
XX
PS Claim 7; Fig 1; 52pp; English.
XX
CC The invention relates to the use of a Fringe protein or a Brainiac
CC protein or a fragment or functional equivalent of a Fringe or a
CC Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC proteins have been found to possess glycosyltransferase activity in
CC transferring sugar residues onto certain proteins of biological
CC interest, so affecting the binding of effector molecules to these
CC proteins. Particularly they are useful for transferring a
CC N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC whether free or attached to a lipid, carbohydrate or protein. They
CC act as glycosyltransferase on notch protein (EGF-module containing
CC protein) family. They are useful in treatment of a disease caused
CC by epidermal growth factor (EGF)-like module containing protein,
CC such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC autosomal dominant arteriopathy with subcortical infarcts,
CC leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC proteins are also useful for screening ligands capable of modulating
CC the activity of the protein. The present sequence is human protein
CC which is a member of Fringe and Brainiac protein family.
XX
SQ Sequence 192 AA;

Query Match 67.8%; Score 1016; DB 23; Length 192;
Best Local Similarity 100.0%; Pred. No. 2.5e-109;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      33 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 92
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 60

Qy      93 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTG 152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTG 120

Qy     153 FLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLE 212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLE 180

Qy     213 DVFVGLCLERLN 224
           ||||||||||||
Db     181 DVFVGLCLERLN 192
```

RESULT 4
AAR57433
ID AAR57433 standard; Protein; 326 AA.
XX
AC AAR57433;
XX
DT 14-MAR-1995 (first entry)
XX
DE Beta-1,3-galactosyltransferase.
XX
KW Beta-1,3-galactosyltransferase; enzyme; saccharide chain; pAMoPRWM1;
KW KJM-1 cells.
XX
OS Homo sapiens.
XX
PN JP06181759-A.
XX
PD 05-JUL-1994.
XX
PF 16-DEC-1992; 92JP-0336436.
XX
PR 16-DEC-1992; 92JP-0336436.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR WPI; 1994-251683/31.
DR N-PSDB; AAR57433.
XX
PT Beta-galactosyl-transferase DNA and protein - useful for prodn. of
PT saccharide chains
XX
PS Claim 1; Page 22-24; 47pp; Japanese.
XX
CC This sequence represents a beta-1,3-galactosyltransferase. This
CC enzyme can be used to produce physiologically active saccharide
CC chains and varients, and for improvement of saccharide chains bound
CC to physiologically active proteins. The cDNA encoding this protein
CC represents a fragment of the plasmid pAMoPRWM1 which was cloned in
CC KJM-1 cells.
XX
SQ Sequence 326 AA;

Query Match 38.5%; Score 577; DB 15; Length 326;
Best Local Similarity 42.8%; Pred. No. 5.5e-58;
Matches 116; Conservative 46; Mismatches 93; Indels 16; Gaps 4;

```
Qy      14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
           ||       | :  |||||:|::::||:   | |||:||| |   || :: | |||| :
Db      65 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 124

Qy      74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
              : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||||:|:|:
Db     125 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 184

Qy     134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
           | |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
Db     185 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 243

Qy     192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
           | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db     244 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 295

Qy     247 LFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
           :||::  | | |  :   |  : : :   |
Db     296 RYRRVITVHQISPEEMHRIWNDMSSKKHLRC 326
```

RESULT 5
AAY02366
ID AAY02366 standard; Protein; 422 AA.
XX
AC AAY02366;
XX
DT 13-JUL-1999 (first entry)
XX
DE Polypeptide identified by the signal sequence trap method.
XX
KW Signal sequence trap method; SST method; immunisation; inhibition;
KW infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW activin activity; inhibin activity; chemokine activity;
KW cytokine activity; blood coagulation regulation; agonist; antagonist;
KW metabolic disorder; hormonal disorder; immune disorder; wound;
KW severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer.

```
XX
OS   Homo sapiens.
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
XX
DR   WPI; 1999-277254/23.
DR   N-PSDB; AAX35710, AAX35711.
XX
PT   Polypeptides identified by the signal sequence trap method from a
PT   human cDNA library
XX
PS   Claim 1; Page 106-108; 281pp; Japanese.
XX
CC   AAY02358-84 represent novel polypeptides which are identified from a
CC   human placental cDNA library by the signal sequence trap (SST) method.
CC   The polypeptides are encoded by the cDNA seqeuences in AAX35694-X35747.
CC   The polypeptides have a broad range of physiological activity, including
CC   immunisation against and inhibition of infections, allergies and cancer;
CC   regulation of tissue formation and repair; activin/inhibin activity;
CC   chemokine/cytokine activity; blood coagulation regulation; and
CC   receptor/ligand agonist or antagonist activity. The polypeptides can
CC   be used for prevention and treatment of disorders including infections
CC   by bacteria, yeasts and viruses (including HIV) and protozoa; metabolic
CC   and hormonal disorders; immune disorders (including severe combined
CC   immunodeficiency (SCID) and AIDS; thrombosis; cancer; and traumatic or
CC   surgical wounds.
XX
SQ   Sequence   422 AA;

  Query Match            38.0%;  Score 569;  DB 20;  Length 422;
  Best Local Similarity  41.5%;  Pred. No. 7e-57;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
            :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
           ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     258 KPRTLLDYWQALENSRGEDC 277
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405


RESULT 6
AAY06461
ID   AAY06461 standard; Protein; 319 AA.
XX
AC   AAY06461;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Human Dendriac.
XX
KW   Dendriac; Brainiac-2; human; Notch; immune disorder;
KW   neurological disorder; diagnosis; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /note= "signal peptide"
FT   Protein         26..319
FT                   /note= "mature protein"
FT   Modified-site   60
FT                   /note= "N-glycosylated"
FT   Modified-site   142
FT                   /note= "N-glycosylated"
FT   Modified-site   186
FT                   /note= "N-glycosylated"
FT   Modified-site   200
FT                   /note= "N-glycosylated"
FT   Modified-site   314
FT                   /note= "N-glycosylated"
FT   Domain          82..90
FT                   /label= CD-I
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          100..109
FT                   /label= CD-II
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          120..153
FT                   /label= CD-III
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          168..182
FT                   /label= CD-IV
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          199..204
FT                   /label= CD-V
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          217..225
FT                   /label= CD-VI
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          234..242
FT                   /label= CD-VII
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          250..261
FT                   /label= CD-VIII
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          274..282
FT                   /label= CD-IX
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          285..291
FT                   /label= CD-X
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          297..308
FT                   /label= CD-XI
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Peptide         55..63
FT                   /note= "epitope-bearing peptide"
FT   Peptide         75..84
FT                   /note= "epitope-bearing peptide"
FT   Peptide         89..95
FT                   /note= "epitope-bearing peptide"
FT   Peptide         233..238
FT                   /note= "epitope-bearing peptide"
```

```
FT   Peptide         249..255
FT                   /note= "epitope-bearing peptide"
FT   Peptide         271..277
FT                   /note= "epitope-bearing peptide"
FT   Peptide         298..302
FT                   /note= "epitope-bearing peptide"
XX
PN   WO9931116-A1.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;    98WO-US27049.
XX
PR   17-NOV-1998;    98US-0108928.
PR   18-DEC-1997;    97US-0068006.
PR   12-MAR-1998;    98US-0077687.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ebner R,  Endress GA,  Florence KA,  Rosen CA,  Ruben SM;
PI   Soppet DR,  Yu G;
XX
DR   WPI; 1999-430031/36.
DR   N-PSDB; AAX87192, AAX87193.
XX
PT   Human proteins for treating and detecting immune and neurological
PT   disorders
XX
PS   Claim 22; Page 112-113; 132pp; English.
XX
CC   The present sequence represents human Dendriac, also called
CC   Brainiac-2, a novel member of the Brainiac family.  The sequence
CC   was deduced from Dendriac cDNA (see AAX87192 and AAX87193) initially
CC   discovered in a dendritic cell library.  Additional clones of the
CC   same gene were identified in NTERA2 cell, adult pulmonary tissue,
CC   salivary gland, ovary, Caco-2 colon adenocarcinoma, smooth muscle,
CC   cerebellum, 8-week-old human mrbyo, haemagiopericytoma, amygdala,
CC   substantia nigra and whole brain cDNA libraries.  Homology is
CC   shown to Brainiac-3 (see AAY06462) and to Drosophila Brainiac.
CC   Recombinant Dendriac polypeptides, especially the mature protein
CC   or epitope-bearing portions of it, can be expressed in host cells
CC   using claimed recombinant vectors.  Dendriac and Brainiac proteins
CC   are members of the Notch family of proteins involved in the control
CC   of cell division.  The proteins may be used to detect and/or treat
CC   immune system and neurological disorders, and to raise specific
CC   antibodies.
XX
SQ   Sequence   319 AA;

  Query Match             36.6%;  Score 548.5;  DB 20;  Length 319;
  Best Local Similarity   38.6%;  Pred. No. 1.1e-54;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      40 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 99

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     100 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 159

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     160 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 219

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     220 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 279

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     280 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 311
```

```
RESULT 7
AAY66685
ID   AAY66685 standard; protein; 331 AA.
XX
AC   AAY66685;
XX
DT   05-APR-2000  (first entry)
XX
DE   Membrane-bound protein PRO1074.
XX
KW   Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW   pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS   Homo sapiens.
XX
PN   WO9963088-A2.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999;    99WO-US12252.
XX
PR   02-JUN-1998;    98US-0087607.
PR   02-JUN-1998;    98US-0087609.
PR   02-JUN-1998;    98US-0087759.
PR   03-JUN-1998;    98US-0087827.
PR   04-JUN-1998;    98US-0088021.
PR   04-JUN-1998;    98US-0088025.
PR   04-JUN-1998;    98US-0088028.
PR   04-JUN-1998;    98US-0088029.
PR   04-JUN-1998;    98US-0088030.
PR   04-JUN-1998;    98US-0088033.
PR   04-JUN-1998;    98US-0088326.
PR   05-JUN-1998;    98US-0088167.
PR   05-JUN-1998;    98US-0088202.
PR   05-JUN-1998;    98US-0088212.
PR   05-JUN-1998;    98US-0088217.
PR   09-JUN-1998;    98US-0088655.
PR   10-JUN-1998;    98US-0088722.
PR   10-JUN-1998;    98US-0088730.
PR   10-JUN-1998;    98US-0088734.
PR   10-JUN-1998;    98US-0088738.
PR   10-JUN-1998;    98US-0088740.
PR   10-JUN-1998;    98US-0088741.
PR   10-JUN-1998;    98US-0088742.
PR   10-JUN-1998;    98US-0088810.
PR   10-JUN-1998;    98US-0088811.
PR   10-JUN-1998;    98US-0088824.
PR   10-JUN-1998;    98US-0088825.
PR   10-JUN-1998;    98US-0088826.
PR   11-JUN-1998;    98US-0088858.
PR   11-JUN-1998;    98US-0088861.
PR   11-JUN-1998;    98US-0088863.
PR   11-JUN-1998;    98US-0088876.
PR   12-JUN-1998;    98US-0089090.
PR   12-JUN-1998;    98US-0089105.
PR   16-JUN-1998;    98US-0089440.
PR   16-JUN-1998;    98US-0089512.
PR   16-JUN-1998;    98US-0089514.
PR   17-JUN-1998;    98US-0089532.
PR   17-JUN-1998;    98US-0089538.
PR   17-JUN-1998;    98US-0089598.
PR   17-JUN-1998;    98US-0089599.
PR   17-JUN-1998;    98US-0089600.
PR   17-JUN-1998;    98US-0089653.
PR   18-JUN-1998;    98US-0089801.
PR   18-JUN-1998;    98US-0089907.
PR   18-JUN-1998;    98US-0089908.
PR   19-JUN-1998;    98US-0089947.
PR   19-JUN-1998;    98US-0089948.
PR   19-JUN-1998;    98US-0089952.
PR   22-JUN-1998;    98US-0090246.
PR   22-JUN-1998;    98US-0090252.
```

PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
DR WPI; 2000-072883/06.
DR N-PSDB; AAZ65022.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
PS claim 12; Fig 137; 822pp; English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
SQ Sequence 331 AA;

Query Match 36.6%; Score 548.5; DB 21; Length 331;
Best Local Similarity 38.6%; Pred. No. 1.2e-54;
Matches 105; Conservative 53; Mismatches 111; Indels 3; Gaps 1;

```
Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323
```

RESULT 8
AAU29104
ID AAU29104 standard; Protein; 331 AA.
XX
AC AAU29104;
XX

DT 18-DEC-2001 (first entry)
XX
DE Human PRO polypeptide sequence #81.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2001-602746/68.
DR N-PSDB; AAS46005.
XX
PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
PS Claim 11; Fig 162; 774pp; English.
XX
CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC The PRO polypeptides and their associated nucleic acids can be used to
CC detect the presence of a tumour in a mammal by comparing the level of
CC expression of a PRO polypeptide in a test sample of cells from the animal
CC and a control sample of normal cells, whereby a higher level of
CC expression in the test sample indicates the presence of a tumour in the
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC and rabbits but are preferably human. The polypeptides can be used to
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC when contacted with it. A specific polypeptide can be used to stimulate
CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
SQ Sequence 331 AA;

Query Match 36.6%; Score 548.5; DB 22; Length 331;
Best Local Similarity 38.6%; Pred. No. 1.2e-54;
Matches 105; Conservative 53; Mismatches 111; Indels 3; Gaps 1;

```
Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323
```

RESULT 9
AAB87543
ID AAB87543 standard; Protein; 331 AA.
XX
AC AAB87543;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1074.
XX
KW Human; PRO protein; mapping.
XX
OS Homo sapiens.
XX
PN WO200116318-A2.
XX
PD 08-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-US23328.
XX
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.

```
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   N-PSDB; AAF92075.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 12; Fig 36; 278pp; English.
XX
CC   The present sequence is a human PRO polypeptide (secreted and
CC   transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC   anti-PRO antibodies are useful for preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the PRO protein,
CC   agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC   employed as molecular weight markers for protein electrophoresis. The PRO
CC   coding sequence has applications in molecular biology, including use as
CC   hybridisation probes, and in chromosome and gene mapping.
XX
SQ   Sequence   331 AA;

  Query Match             36.6%;  Score 548.5;  DB 22;  Length 331;
  Best Local Similarity   38.6%;  Pred. No. 1.2e-54;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 10
AAB65208
ID   AAB65208 standard; Protein; 331 AA.
XX
AC   AAB65208;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO1074 (UNQ531) protein sequence SEQ ID NO:209.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   N-PSDB; AAF44168.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 12; Fig 137; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence   331 AA;

  Query Match             36.6%;  Score 548.5;  DB 22;  Length 331;
  Best Local Similarity   38.6%;  Pred. No. 1.2e-54;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171
```

```
Qy    118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
          | :|||||:|:|||   | : ||  | : :||||:  :: :  |  : |  :|  |||:
Db    172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy    178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
            :||:||| ||: | |:  ::| :   |  || |||:||:||  | : :         ||
Db    232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy    238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
            :   ||  ||::| |    : :: :|| :
Db    292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 11
AAU80224
ID   AAU80224 standard; Protein; 331 AA.
XX
AC   AAU80224;
XX
DT   30-JUL-2002  (first entry)
XX
DE   Human beta1,3-acetylgalactosamine transferase.
XX
KW   Human; enzyme; beta1,3-acetylgalactosamine transferase;
KW   Gb4 sugar chain.
XX
OS   Homo sapiens.
XX
PN   JP2002085069-A.
XX
PD   26-MAR-2002.
XX
PF   08-SEP-2000; 2000JP-0273835.
XX
PR   08-SEP-2000; 2000JP-0273835.
XX
PA   (SEGK ) SEIKAGAKU KOGYO CO LTD.
PA   (FURU/) FURUKAWA K.
XX
DR   WPI; 2002-378274/41.
DR   N-PSDB; ABK51201.
XX
PT   Preparation of beta-1,3-acetylgalactosamine transferase -
XX
PS   Claim 3; Page 13-14; 15pp; Japanese.
XX
CC   The invention relates to the preparation of beta-1,3-acetylgalactosamine
CC   (a Gb4 sugar chain molecule) comprising introducing a DNA appearing as
CC   ABK51201 (S1) or a DNA hybridising with S1 or a base sequence
CC   complementary to the base sequence or part of these base sequences under
CC   a stringent condition into a cell and growing the cell to express beta-1,
CC   3-acetylgalactosamine transferase and collecting it. Also included is the
CC   preparation of Gb4 sugar chain comprising contacting a polypeptide
CC   comprising the beta-1,3-acetylgalactosamine transferase protein appearing
CC   as AAU80224 (A) or a polypeptide consisting of an amino acid sequence in
CC   which at least one amino acid is replaced, deleted, inserted or
CC   transferred in the amino acid sequence (A) and having enzymatic activity
CC   transferring N-acetylgalactosamine residue from an N-acetylgalactosamine
CC   donor to the C3 site of the galactose residue in Gb3 sugar chain which is
CC   the receptor with an N-acetylgalactosamine donor and Gb3 sugar chain.
CC   The present sequence represents beta-1,3-acetylgalactosamine transferase.
XX
SQ   Sequence   331 AA;

  Query Match             36.6%;  Score 548.5;  DB 7;  Length 331;
  Best Local Similarity   38.6%;  Pred. No. 1.2e-54;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy      1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
            | |   :|::| :|     ::|    ||||:||||    :  | ||| |||:::   | :
Db     52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy     61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
          : ||||||  :   :      :: |   :||||::|||| | ||||||:|   ||  |||
Db    112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy    118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
          | :|||||:|:|||   | : ||  | : :||||:  :: :  |  : |  :|  |||:
Db    172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy    178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
            :||:||| ||: | |:  ::| :   |  || |||:||:||  | : :         ||
Db    232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy    238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
            :   ||  ||::| |    : :: :|| :
Db    292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 12
AAB93992
ID   AAB93992 standard; Protein; 331 AA.
XX
AC   AAB93992;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14093.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14093; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
```

```
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   331 AA;

  Query Match             36.3%;  Score 543.5;  DB 22;  Length 331;
  Best Local Similarity   38.6%;  Pred. No. 4.4e-54;
  Matches  105;  Conservative   52;  Mismatches  112;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     232 KVSPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 13
AAE15927
ID   AAE15927 standard; Protein; 193 AA.
XX
AC   AAE15927;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human 4566767/84-303 protein, member of Fringe and Brainiac family.
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; human.
XX
OS   Homo sapiens.
XX
PN   WO200187321-A2.
XX
PD   22-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-IB01033.
XX
PR   19-MAY-2000; 2000GB-0012216.
XX
PA   (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Cohen S,  Brueckner K,  Clausen H,  Keck B;
XX
DR   WPI; 2002-082940/11.
XX
PT   Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT   treatment of disease caused by epidermal growth factor-like module
PT   containing protein, such as T cell leukemia, breast cancer, stroke and
PT   dementia  -
XX
PS   Claim 7; Fig 1; 52pp; English.
XX
CC   The invention relates to the use of a Fringe protein or a Brainiac
CC   protein or a fragment or functional equivalent of a Fringe or a
CC   Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC   proteins have been found to possess glycosyltransferase activity in
CC   transferring sugar residues onto certain proteins of biological
CC   interest, so affecting the binding of effector molecules to these
CC   proteins. Particularly they are useful for transferring a
CC   N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC   whether free or attached to a lipid, carbohydrate or protein. They
CC   act as glycosyltransferase on notch protein (EGF-module containing
CC   protein) family. They are useful in treatment of a disease caused
CC   by epidermal growth factor (EGF)-like module containing protein,
CC   such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC   autosomal dominant arteriopathy with subcortical infarcts,
CC   leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC   proteins are also useful for screening ligands capable of modulating
CC   the activity of the protein. The present sequence is human protein
CC   which is a member of Fringe and Brainiac protein family.
XX
SQ   Sequence   193 AA;

  Query Match             33.2%;  Score 497.5;  DB 23;  Length 193;
  Best Local Similarity   49.7%;  Pred. No. 4.3e-49;
  Matches   96;  Conservative   34;  Mismatches   60;  Indels    3;  Gaps    2;

Qy      33 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 92
Db       1 ISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVLNQMVEQESQIFHDIIVE 60

Qy      93 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNR--TTRFF 150
Db      61 DFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDNLIYRLLKPSTKPRRRYF 120

Qy     151 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 210
Db     121 TGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLH 179

Qy     211 LEDVFVGLCLERL 223
Db     180 LEDVYVGLCLRKL 192


RESULT 14
AAE15928
ID   AAE15928 standard; Protein; 193 AA.
XX
AC   AAE15928;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Mouse 2745735/84-303 protein, member of Fringe and Brainiac family.
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; mouse.
XX
OS   Mus sp.
XX
PN   WO200187321-A2.
XX
PD   22-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-IB01033.
XX
PR   19-MAY-2000; 2000GB-0012216.
XX
PA   (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Cohen S,  Brueckner K,  Clausen H,  Keck B;
XX
DR   WPI; 2002-082940/11.
XX
PT   Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT   treatment of disease caused by epidermal growth factor-like module
```

```
PT   containing protein, such as T cell leukemia, breast cancer, stroke and
PT   dementia -
XX
PS   Claim 7; Fig 1; 52pp; English.
XX
CC   The invention relates to the use of a Fringe protein or a Brainiac
CC   protein or a fragment or functional equivalent of a Fringe or a
CC   Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC   proteins have been found to possess glycosyltransferase activity in
CC   transferring sugar residues onto certain proteins of biological
CC   interest, so affecting the binding of effector molecules to these
CC   proteins. Particularly they are useful for transferring a
CC   N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC   whether free or attached to a lipid, carbohydrate or protein. They
CC   act as glycosyltransferase on notch protein (EGF-module containing
CC   protein) family. They are useful in treatment of a disease caused
CC   by epidermal growth factor (EGF)-like module containing protein,
CC   such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC   autosomal dominant arteriopathy with subcortical infarcts,
CC   leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC   proteins are also useful for screening ligands capable of modulating
CC   the activity of the protein. The present sequence is mouse protein
CC   which is a member of Fringe and Brainiac protein family.
XX
SQ   Sequence   193 AA;

  Query Match             33.2%;  Score 497.5;  DB 23;  Length 193;
  Best Local Similarity   49.7%;  Pred. No. 4.3e-49;
  Matches   96;  Conservative   34;  Mismatches   60;  Indels    3;  Gaps    2;

Qy      33 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 92
           ::::||:   | |||:||| |   || :: | |||| :     : |:|||||   ||| :
Db       1 ISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVLNQMVEQESQIFHDIIVE 60

Qy      93 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNR--TTRFF 150
           ||:| |:|||||:||: ||  || :| :|||||||:|:|:| |   ||| :     |:|
Db      61 DFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDNLIYRLLKPSTKPRRRYF 120

Qy     151 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 210
           ||:: :|  |||   |||:: :  ||   |||||||||:|| |||  :|  |      :
Db     121 TGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLH 179

Qy     211 LEDVFVGLCLERL 223
           ||||:|||||  :|
Db     180 LEDVYVGLCLRKL 192


RESULT 15
AAE15929
ID   AAE15929 standard; Protein; 196 AA.
XX
AC   AAE15929;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human 4502339_JBC_1/157-364 protein member of Fringe and Brainiac family.
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; human.
XX
OS   Homo sapiens.
XX
PN   WO200187321-A2.
XX
PD   22-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-IB01033.
XX
PR   19-MAY-2000; 2000GB-0012216.
XX
PA   (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Cohen S,  Brueckner K,  Clausen H,  Keck B;
XX
DR   WPI; 2002-082940/11.
XX
PT   Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT   treatment of disease caused by epidermal growth factor-like module
PT   containing protein, such as T cell leukemia, breast cancer, stroke and
PT   dementia -
XX
PS   Claim 7; Fig 1; 52pp; English.
XX
CC   The invention relates to the use of a Fringe protein or a Brainiac
CC   protein or a fragment or functional equivalent of a Fringe or a
CC   Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC   proteins have been found to possess glycosyltransferase activity in
CC   transferring sugar residues onto certain proteins of biological
CC   interest, so affecting the binding of effector molecules to these
CC   proteins. Particularly they are useful for transferring a
CC   N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC   whether free or attached to a lipid, carbohydrate or protein. They
CC   act as glycosyltransferase on notch protein (EGF-module containing
CC   protein) family. They are useful in treatment of a disease caused
CC   by epidermal growth factor (EGF)-like module containing protein,
CC   such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC   autosomal dominant arteriopathy with subcortical infarcts,
CC   leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC   proteins are also useful for screening ligands capable of modulating
CC   the activity of the protein. The present sequence is human protein
CC   which is a member of Fringe and Brainiac protein family.
XX
SQ   Sequence   196 AA;

  Query Match             32.0%;  Score 480;  DB 23;  Length 196;
  Best Local Similarity   48.1%;  Pred. No. 4.8e-47;
  Matches   91;  Conservative   36;  Mismatches   58;  Indels    4;  Gaps    2;

Qy      39 QLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEVDQESQRHGDIIQKDFLD 96
           |:  | ||||||| | :  | |:   |||| :   :    : : :||::: ||||:::||
Db       7 QIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAILEESRQYHDIIQQEYLD 66

Qy      97 VYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTR--FFTGFL 154
           |||||:||:||: ||  :||   :|||||||||:| :||   ||| :   |  :|||:|
Db      67 TYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINRLLKPDLPPRHNYFTGYL 126

Qy     155 KLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDV 214
               | |   |||::    || :||| |||||||||||:| ::: ||  :  : ||||
Db     127 MRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAEKIFKVSLGIRRLHLEDV 186

Qy     215 FVGLCLERL 223
           :||:|| :|
Db     187 YVGICLAKL 195


Search completed: April  4, 2003, 14:44:08
Job time : 40 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:40:30 ; Search time 11 Seconds
(without alignments)
1055.761 Million cell updates/sec

Title: US-09-914-152-1_COPY_31_310
Perfect score: 1498
Sequence: 1 FKEQSFVYKKDGNFLKLPDT.....TLLDYWQALENSRGEDCPPV 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1498	100.0	310	1	B3G5_HUMAN	Q9y2c3 h beta-1,3-
2	1433	95.7	300	1	B3G5_GORGO	Q9n293 g beta-1,3-
3	1422	94.9	301	1	B3G5_PANPA	Q9n294 p beta-1,3-
4	1407	93.9	297	1	B3G5_PANTR	Q9n295 p beta-1,3-
5	1133.5	75.7	308	1	B3G5_MOUSE	Q9ji67 m beta-1,3-
6	413	27.6	372	1	B3G8_HUMAN	Q9y2a9 h probable
7	360	24.0	398	1	B3G7_HUMAN	Q9ny97 h beta-1,3-
8	105.5	7.0	1049	1	VP39_YEAST	Q07468 saccharomyc
9	93.5	6.2	336	1	MTHB_HAEPH	P50193 haemophilus
10	90.5	6.0	1314	1	SS22_YEAST	P25390 saccharomyc
11	89.5	6.0	786	1	TLR1_HUMAN	Q15399 homo sapien
12	88.5	5.9	579	1	YHV0_YEAST	P38848 saccharomyc
13	86.5	5.8	516	1	YVNF_AZOCH	P24423 azotobacter
14	85	5.7	1073	1	HSER_PIG	P55204 sus scrofa
15	84.5	5.6	711	1	ETF2_MYXVL	Q9q8k4 myxoma viru
16	84	5.6	306	1	YFV9_SCHPO	Q9p7q6 schizosacch
17	83	5.5	275	1	ISPH_THEMA	Q9x1f7 thermotoga
18	82.5	5.5	525	1	DIMH_CAEEL	O17397 caenorhabdi
19	82.5	5.5	713	1	PMIP_HUMAN	Q99797 homo sapien
20	82	5.5	576	1	HXT4_YEAST	P32467 saccharomyc
21	81.5	5.4	359	1	UNG_YEAST	P12887 saccharomyc
22	81	5.4	397	1	O22A_DROME	P81909 drosophila
23	81	5.4	1042	1	SYI_BORBU	O51773 borrelia bu
24	80.5	5.4	3386	1	POLG_DEN4	P09866 d genome po
25	79.5	5.3	612	1	BIR3_MOUSE	Q62210 mus musculu
26	79.5	5.3	691	1	YIO4_YEAST	P40460 saccharomyc
27	78.5	5.2	757	1	METE_PASMU	P57843 pasteurella
28	78.5	5.2	1002	1	YA37_ANASP	P58612 anabaena sp
29	78.5	5.2	4377	1	ANK3_HUMAN	Q12955 homo sapien
30	78	5.2	469	1	LEU2_BUCAP	O85065 buchnera ap
31	78	5.2	519	1	YG31_METJA	Q59025 methanococc
32	78	5.2	618	1	BIR3_HUMAN	Q13490 homo sapien
33	77.5	5.2	271	1	Y398_METJA	Q57841 methanococc
34	77.5	5.2	710	1	PMIP_RAT	Q01992 rattus norv
35	77	5.1	246	1	Y402_BUCAI	P57482 buchnera ap
36	77	5.1	741	1	GNT5_HUMAN	Q09328 homo sapien
37	76.5	5.1	397	1	OM1C_CHLTR	P08780 chlamydia t
38	76.5	5.1	711	1	ETF2_SFVKA	Q9q8y2 shope fibro
39	76.5	5.1	977	1	KIT_BOVIN	P43481 bos taurus
40	76	5.1	374	1	GLY1_CANAL	O13427 candida alb
41	76	5.1	2077	1	TEGU_HSV6U	P52340 human herpe
42	75	5.0	333	1	Y642_METJA	Q58059 methanococc
43	75	5.0	454	1	YE62_HAEIN	P45217 haemophilus
44	75	5.0	467	1	XANB_XANCP	P29956 x xanthan b
45	75	5.0	494	1	DCE2_ARATH	Q42472 arabidopsis

ALIGNMENTS

RESULT 1
B3G5_HUMAN
ID B3G5_HUMAN STANDARD; PRT; 310 AA.
AC Q9Y2C3; Q9NY96; Q9P1X7; Q9P1X6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5).
GN B3GALT5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal adenocarcinoma;
RX MEDLINE=99230269; PubMed=10212226;
RA Isshiki S., Togayachi A., Kudo T., Nishihara S., Watanabe M.,
RA Kubota T., Kitajima M., Shiraishi N., Sasaki K., Andoh T.,
RA Narimatsu H.;
RT "Cloning, expression, and characterization of a novel UDP-
RT galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
RT (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal
RT and pancreatic epithelia and tumor cells derived therefrom.";
RL J. Biol. Chem. 274:12499-12507(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99337698; PubMed=10406968;
RA Zhou D., Berger E.G., Hennet T.;
RT "Molecular cloning of a human UDP-galactose:GlcNAcbeta1,3GalNAc
RT beta1, 3 galactosyltransferase gene encoding an O-linked
RT core3-elongation enzyme.";
RL Eur. J. Biochem. 263:571-576(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Amado M., Carneiro F., Clausen H.;
RT "Cloning and expression of two beta-1,3-galactosyltransferases:
RT beta3gal-T5 and beta3gal-T6.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,

```
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [5]
RP   SEQUENCE OF 1-298 FROM N.A.
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   REVIEW.
RX   MEDLINE=20047730; PubMed=10580128;
RA   Amado M., Almeida R., Schwientek T., Clausen H.;
RT   "Identification and characterization of large galactosyltransferase
RT   gene families: galactosyltransferases for all functions.";
RL   Biochim. Biophys. Acta 1473:35-53(1999).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR.
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN STOMACH, JEJUNUM, COLON,
CC       PANCREAS, SMALL INTESTINE, TESTIS AND GASTROINTESTINAL AND
CC       PANCREATIC CANCER CELL LINES. HARDLY DETECTED IN LUNG, LIVER,
CC       ADRENAL GLAND AND PERIPHERAL BLOOD LEUKOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB020337; BAA77664.1; -.
DR   EMBL; AF145784; AAF07880.1; -.
DR   EMBL; AJ006078; CAB91547.1; -.
DR   EMBL; AL163280; CAB90446.1; -.
DR   EMBL; AB041412; BAA94497.1; -.
DR   EMBL; AB041413; BAA94498.1; -.
DR   Genew; HGNC:920; B3GALT5.
DR   MIM; 604066; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    310       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    231    231       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     23     23       F -> Y (IN REF. 3).
FT   CONFLICT     26     26       Y -> N (IN REF. 3).
FT   CONFLICT     42     42       G -> V (IN REF. 5; BAA94497).
FT   CONFLICT     85     85       M -> T (IN REF. 3 AND 5; BAA94497).
FT   CONFLICT    114    114       Q -> R (IN REF. 5; BAA94497).
FT   CONFLICT    234    234       K -> E (IN REF. 5; BAA94497).
FT   CONFLICT    277    277       L -> R (IN REF. 5; BAA94497).
FT   CONFLICT    293    297       LDYWQ -> WTTGR (IN REF. 5; BAA94498).
SQ   SEQUENCE   310 AA;  36189 MW;  4DD7A19E3E648AA9 CRC64;

  Query Match             100.0%;  Score 1498;  DB 1;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-135;
  Matches  280;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 90

Qy      61 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 150

Qy     121 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 210

Qy     181 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270

Qy     241 LRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV 280
           |||||||||||||||||||||||||||||||||||||||||
Db     271 LRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV 310


RESULT 2
B3G5_GORGO
ID   B3G5_GORGO     STANDARD;      PRT;   300 AA.
AC   Q9N293;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE   (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN   B3GALT5.
OS   Gorilla gorilla gorilla (Lowland gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX   NCBI_TaxID=9595;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR (BY SIMILARITY).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB041416; BAA94501.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29   >300       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    130    130       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    231    231       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     300    300
SQ   SEQUENCE   300 AA;  35104 MW;  1AA11692ED9F06FA CRC64;

  Query Match             95.7%;  Score 1433;  DB 1;  Length 300;
  Best Local Similarity   99.6%;  Pred. No. 3.8e-129;
  Matches  269;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      31 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQ 90
```

```
Qy      61 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 150

Qy     121 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 210

Qy     181 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270

Qy     241 LRFSVCLFRRIVACHFIKPRTLLDYWQALE 270
           |||||||||||||||||||||||||||||
Db     271 LRFSVCLFRRIVACHFIKPRTLLDYWQALE 300
```

RESULT 3
B3G5_PANPA
ID B3G5_PANPA STANDARD; PRT; 301 AA.
AC Q9N294;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN B3GALT5.
OS Pan paniscus (Pygmy chimpanzee) (Bonobo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9597;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AB041415; BAA94500.1; -.
DR InterPro; IPR002659; Galactosyl_T.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 >301 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 301 301
SQ SEQUENCE 301 AA; 35222 MW; E04CC70F41CB29FD CRC64;

 Query Match 94.9%; Score 1422; DB 1; Length 301;
 Best Local Similarity 98.5%; Pred. No. 4.3e-128;
 Matches 267; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      31 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQ 90

Qy      61 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 120
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      91 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDGYYNLTLKTMMGIEWVHRFCPQAA 150

Qy     121 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 210

Qy     181 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
           |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db     211 PPFCSGTGYVFSGDVASQVYNVSESVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270

Qy     241 LRFSVCLFRRIVACHFIKPRTLLDYWQALEN 271
           |||||| |||||||||||||||||||||||
Db     271 LRFSVCRFRRIVACHFIKPRTLLDYWQALEN 301
```

RESULT 4
B3G5_PANTR
ID B3G5_PANTR STANDARD; PRT; 297 AA.
AC Q9N295;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE 1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE galactosyltransferase 5) (Beta-3-Gx-T5) (Fragment).
GN B3GALT5.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC EFFICIENT ACCEPTOR (BY SIMILARITY).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AB041414; BAA94499.1; -.
DR InterPro; IPR002659; Galactosyl_T.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Multigene family.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 >297 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 297 297
SQ SEQUENCE 297 AA; 34850 MW; BB7963250A837A28 CRC64;

 Query Match 93.9%; Score 1407; DB 1; Length 297;

```
  Best Local Similarity   98.5%;  Pred. No. 1.1e-126;
  Matches  263;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           ||||||||||||||||||||||||||||||||||:||||||||||||||| |||||
Db      31 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHRQLAERMAIRQTWGKERTVKGKQ 90

Qy      61 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 LKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA 150

Qy     121 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRY 210

Qy     181 PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
           ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db     211 PPFCSGTGYVFSGDVASQVYNVSESVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 270

Qy     241 LRFSVCLFRRIVACHFIKPRTLLDYWQ 267
           |||||| ||||||||||||||||||||
Db     271 LRFSVCRFRRIVACHFIKPRTLLDYWQ 297


RESULT 5
B3G5_MOUSE
ID   B3G5_MOUSE     STANDARD;      PRT;   308 AA.
AC   Q9JI67;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 5 (EC 2.4.1.-) (Beta-1,3-GalTase 5)
DE   (Beta3Gal-T5) (b3Gal-T5) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 5) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 5) (Beta-3-Gx-T5) (Stage-specific embryonic
DE   antigen-3 synthase) (SSEA-3 synthase).
GN   B3GALT5 OR B3GT5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RX   MEDLINE=20390006; PubMed=10837462;
RA   Zhou D., Henion T.R., Jungalwala F.B., Berger E.G., Hennet T.;
RT   "The beta1,3-galactosyltransferase beta3GalT-V is a stage-specific
RT   embryonic antigen-3 (SSEA-3) synthase.";
RL   J. Biol. Chem. 275:22631-22634(2000).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF GAL TO GLCNAC-BASED ACCEPTORS
CC       WITH A PREFERENCE FOR THE CORE3 O-LINKED GLYCAN
CC       GLCNAC(BETA1,3)GALNAC STRUCTURE. CAN USE GLYCOLIPID LC3CER AS AN
CC       EFFICIENT ACCEPTOR. ALSO CATALYZES THE TRANSFER OF GAL TO THE
CC       TERMINAL GALNAC UNIT OF THE GLOBOSIDE GB4, THEREBY SYNTHESIZING
CC       THE GLYCOLIPID GB5, ALSO KNOWN AS THE STAGE-SPECIFIC EMBRYONIC
CC       ANTIGEN-3 (SSEA-3).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF254738; AAF86241.1; -.
DR   MGD; MGI:2136878; B3galt5.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     25       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       26    308       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    172    172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   308 AA;  35964 MW;  789073A5178825B1 CRC64;

  Query Match             75.7%;  Score 1133.5;  DB 1;  Length 308;
  Best Local Similarity   73.3%;  Pred. No. 1.4e-100;
  Matches  206;  Conservative   36;  Mismatches   38;  Indels    1;  Gaps    1;

Qy       1 FKEQSFVYKKD-GNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGK 59
           |:|  ||:||  | ||::|| ||:| |||||||||||||||| |||||:|||:|  |:|:
Db      28 FRELPFVFKKSHGKFLQIPDIDCKQKPPFLVLLVTSSHKQLAARMAIRKTWGRETSVQGQ 87

Qy      60 QLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQA 119
           |::||||||: |  |      ||::| |||||| | |:|||||||||:|||: ||||
Db      88 QVRTFFLLGTSDSTEEMDATTLESEQHRDIIQKDFKDAYFNLTLKTMMGMEWVYHFCPQT 147

Qy     120 AFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDR 179
           |:|||||||||:|| |||||||||:|||||||::| ::||||| |:|||||| ||||||
Db     148 AYVMKTDSDMFVNVGYLTELLLKKNKTTRFFTGYIKPHDFPIRQKFNKWFVSKFEYPWDR 207

Qy     180 YPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPG 239
           |||||||||||||| ||| |||||:|||:|||||||||||| :| || ||||:: |||||
Db     208 YPPFCSGTGYVFSSDVAIQVYNVSESVPFIKLEDVFVGLCLAKLKIRPEELHTKQTFFPG 267

Qy     240 GLRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV 280
           ||||||| |::|||||:||: || ||||||||: :||| |
Db     268 GLRFSVCRFQKIVACHFMKPQDLLTYWQALENSKEQDCPAV 308


RESULT 6
B3G8_HUMAN
ID   B3G8_HUMAN     STANDARD;      PRT;   372 AA.
AC   Q9Y2A9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable beta-1,3-galactosyltransferase 8 (EC 2.4.1.-) (Beta-1,3-
DE   GalTase 8) (Beta3Gal-T8) (b3Gal-T8) (UDP-galactose:beta-N-
DE   acetylglucosamine beta-1,3-galactosyltransferase 8) (UDP-Gal:beta-
DE   GlcNAc beta-1,3-galactosyltransferase 8) (Beta-3-Gx-T8).
GN   B3GALT8 OR TMEM3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99173880; PubMed=10072769;
RA   Yokoyama-Kobayashi M., Yamaguchi T., Sekine S., Kato S.;
RT   "Selection of cDNAs encoding putative type II membrane proteins on the
RT   cell surface from a human full-length cDNA bank.";
RL   Gene 228:161-167(1999).
RN   [2]
RP   REVIEW.
RX   MEDLINE=20047730; PubMed=10580128;
RA   Amado M., Almeida R., Schwientek T., Clausen H.;
RT   "Identification and characterization of large galactosyltransferase
RT   gene families: galactosyltransferases for all functions.";
RL   Biochim. Biophys. Acta 1473:35-53(1999).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB015630; BAA76497.1; -.
DR   Genew; HGNC:13528; B3GNT3.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family.
FT   DOMAIN        1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     11     31       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       32    372       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    202    202       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    362    362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   372 AA;  42515 MW;  A6E3FE88AFCCFCC2 CRC64;

  Query Match             27.6%;  Score 413;  DB 1;  Length 372;
  Best Local Similarity   34.9%;  Pred. No. 7.7e-32;
  Matches   94;  Conservative   52;  Mismatches   99;  Indels   24;  Gaps    6;

Qy      17 LPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAET 76
           :| : | | | ||:|:: ||       |  :|:|||:|| |:| ||:  ||:|| |:  |
Db      98 VPPSKCAQ-PVFLLLVIKSSPSNYVRRELLRRTWGRERKVRGLQLRLLFLVGTASNPHEA 156

Qy      77 KEVDQ----ESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFIN 132
           ::|::    |:| ||||:| || | ::||||| :: ::|    |  |:||:  | |:| :
Db     157 RKVNRLLELEAQTHGDILQWDFHDSFFNLTLKQVLFLQWQETRCANASFVLNGDDDVFAH 216

Qy     133 VDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFS 192
            | :   |   :     | | |  |  |||  :||::| :     :||||:| | |:: |
Db     217 TDNMVFYLQDHDPGRHLFVGQLIQNVGPIRAFWSKYYVPEVVTQNERYPPYCGGGGFLLS 276

Qy     193 GDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPG----GLR------ 242
             |: :   :  :    ::|||:|:|||     || |  :|    |    |:|
Db     277 RFTAAALRRAAHVLDIFPIDDVFLGMCLE-----LEGL--KPASHSGIRTSGVRAPSQHL 329

Qy     243 --FSVCLFRRIVACHFIKPRTLLDYWQAL 269
             |  | :| ::  |   |  :|  | ||
Db     330 SSFDPCFYRDLLLVHRFLPYEMLLMWDAL 358


RESULT 7
B3G7_HUMAN
ID   B3G7_HUMAN     STANDARD;      PRT;   398 AA.
AC   Q9NY97; Q9NQR0; Q9NUT9; Q9NQQ9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-1,3-galactosyltransferase 7 (EC 2.4.1.-) (Beta-1,3-GalTase 7)
DE   (Beta3Gal-T7) (b3Gal-T7) (UDP-galactose:beta-N-acetylglucosamine beta-
DE   1,3-galactosyltransferase 7) (UDP-Gal:beta-GlcNAc beta-1,3-
DE   galactosyltransferase 7) (Beta-3-Gx-T7).
GN   B3GALT7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Amado M., Carneiro F., Clausen H.;
RT   "Cloning and expression of two beta-1,3-galactosyltransferases:
RT   beta3gal-T5 and beta3gal-T6.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Urinary bladder;
RA   Gromova I., Gromov P., Celis J.E.;
RT   "A novel member of beta-1,3-galactosyltransferase family is down
RT   regulated during bladder TCC progression.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 25-398 FROM N.A.
RC   TISSUE=Placenta;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA   Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA   Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA   Masuho Y., Kanehori K.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVIEW.
RX   MEDLINE=20047730; PubMed=10580128;
RA   Amado M., Almeida R., Schwientek T., Clausen H.;
RT   "Identification and characterization of large galactosyltransferase
RT   gene families: galactosyltransferases for all functions.";
RL   Biochim. Biophys. Acta 1473:35-53(1999).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 31.
CC   -!- CAUTION: WAS INDICATED AS B3GAL-T6 IN SUBMITTED DNA ENTRIES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ006077; CAB91546.1; -.
DR   EMBL; AF288208; AAF97253.1; -.
DR   EMBL; AF288209; AAF97254.1; -.
DR   EMBL; AK002009; BAA92031.1; ALT_INIT.
DR   Genew; HGNC:15629; B3GNT1.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack; Multigene family; Alternative splicing.
FT   DOMAIN        1      7       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      8     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    398       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    220    220       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     12       MSVGRRRIKLLL -> MVSRSLV (IN ISOFORM 2).
FT   CONFLICT     12     12       MISSING (IN REF. 2).
SQ   SEQUENCE   398 AA;  46135 MW;  115B829FEB51D2F8 CRC64;

  Query Match             24.0%;  Score 360;  DB 1;  Length 398;
  Best Local Similarity   29.1%;  Pred. No. 9.4e-27;
  Matches   82;  Conservative   66;  Mismatches  120;  Indels   14;  Gaps    7;

Qy       1 FKEQSFVYKKDGNFLKL---PDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVK 57
           ||:   :| :  |:  |   ||   ::  |||:| : |     | | |||::||:|
Db     117 FKD-FLLYLRCRNYSLLIDQPDKCAKK--PFLLLAIKSLTPHFARRQAIRESWGQESNAG 173

Qy      58 GKQLKTFFLLGTTSSAAETKEVDQ----ESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVH 113
           : :   |||| |        ::      ||::| ||:  :: | ::|||:|| :: : ||
Db     174 NQTVVRVFLLGQTPPEDNHPDLSDMLKFESEKHQDILMWNYRDTFFNLSLKEVLFLRWVS 233

Qy     114 RFCPQAAFVMKTDSDMFINVDYLTELL--LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVS 171
             ||   || | | |:|:|  ::   |  | | :    | | :  |  | |    |:::
Db     234 TSCPDTEFVFKGDDDVFVNTHHILNYLNSLSKTKAKDLFIGDVIHNAGPHRDKKLKYYIP 293
```

```
Qy     172 KSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELH 231
Db     294 EVVYS-GLYPPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLQKLGLVPEKHK 352

Qy     232 SQPTF-FPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALENS 272
Db     353 GFRTFDIEEKNKNNICSYVDLMLVHSRKPQEMIDIWSQLQSA 394


RESULT 8
VP39_YEAST
ID   VP39_YEAST     STANDARD;      PRT;  1049 AA.
AC   Q07468;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Vacuolar assembly protein VPS39 (Vacuolar morphogenesis protein VAM6).
GN   VPS39 OR VAM6 OR YDL077C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97269041; PubMed=9111041;
RA   Nakamura N., Hirata A., Ohsumi Y., Wada Y.;
RT   "Vam2/Vps41p and Vam6/Vps39p are components of a protein complex on
RT   the vacuolar membranes and involved in the vacuolar assembly in the
RT   yeast Saccharomyces cerevisiae.";
RL   J. Biol. Chem. 272:11344-11349(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Wambutt R., Wedler H., Wedler E., Scharfe M.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR VACUOLAR ASSEMBLY.
CC   -!- SUBUNIT: INTERACTS WITH VPS41/VAM2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74125; CAA98643.1; -.
DR   EMBL; D83058; BAA11758.1; -.
DR   SGD; S0002235; VAM6.
SQ   SEQUENCE   1049 AA;  122881 MW;  EE8D94E7E2F5F078 CRC64;

  Query Match             7.0%;  Score 105.5;  DB 1;  Length 1049;
  Best Local Similarity   24.4%;  Pred. No. 0.057;
  Matches   44;  Conservative   32;  Mismatches   53;  Indels   51;  Gaps    9;

Qy       7 VYKKDGNFLKL----PD--------------------------TDCRQTPPFLVLLVTS 35
Db      38 VYSRDGNTLKLFQVYPDLLQNAKNDPLPPVIENFYFANELSTIFAQCKET---LILLSTT 94

Qy      36 S---HKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQ- 91
Db      95 NLHEYDRIIDRRGINHCWLFERSHKNKEEKNTYLIYSTINTAKMRVLIWEGRTYKNMMEA 154

Qy      92 -----KDFLDVYY----NLTLKTMMGI-EWVHRFCPQAAFVMKTDSDMFIN--VDYLTEL 139
Db     155 SLSYRKETIRSIYPGETGITLATDLGIYHWPYN-KPSLIRIEKTVKNKFPKDMISALTEL 213


RESULT 9
MTHB_HAEPH
ID   MTHB_HAEPH     STANDARD;      PRT;   336 AA.
AC   P50193;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Modification methylase HphIB (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase HphIB) (M.HphIB) (M.HphI(A)).
GN   HPHIBM OR HPHIMA.
OS   Haemophilus parahaemolyticus.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=735;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49700;
RX   MEDLINE=96313243; PubMed=8759008;
RA   Lubys A., Lubiene J., Kulakauskkas S., Stankevicius K.,
RA   Timinskas A., Janulaitis A.;
RT   "Cloning and analysis of the genes encoding the type IIS restriction-
RT   modification system HphI from Haemophilus parahaemolyticus.";
RL   Nucleic Acids Res. 24:2760-2766(1996).
CC   -!- FUNCTION: THIS METHYLASE PROTECTS THE DNA FROM CLEAVAGE BY THE
CC       HPHI ENDONUCLEASE.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85374; CAA59691.1; -.
DR   REBASE; 3661; M.HphIB.
DR   InterPro; IPR002294; D12N6_mtfrase.
DR   InterPro; IPR002052; N6_Mtase.
DR   Pfam; PF02086; MethyltransfD12; 1.
DR   PRINTS; PR00505; D12N6MTFRASE.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Transferase; Methyltransferase; Restriction system.
SQ   SEQUENCE   336 AA;  39609 MW;  3BD144C7772C4FDF CRC64;

  Query Match             6.2%;  Score 93.5;  DB 1;  Length 336;
  Best Local Similarity   22.3%;  Pred. No. 0.19;
  Matches   39;  Conservative   27;  Mismatches   58;  Indels   51;  Gaps    7;

Qy      82 ESQRHG-DIIQKDFLDVYYNLTLK---------TMMGIEWVHRFCPQAAFVMKTDSDMFI 131
Db      46 EAKKRGYQVIANDILNINYQLALALIVNNQEILTACDVDFIFSNPPKSGFMTKNYSDVFF 105

Qy     132 ------NVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQ-PFSKWFVS--------KSEYP 176
Db     106 FKEECRELDAIRANILKLNNTYKQALAFALMRRAMIRKMPYSRFTISWEKVKQLRDEEYS 165

Qy     177 WDRY----------------PPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVF 215
Db     166 YSKYGRRRAYHNQSFEFHFRENLNSYNQAVFNNGNIHQAYN----------EDVF 210


RESULT 10
SS22_YEAST
ID   SS22_YEAST     STANDARD;      PRT;  1314 AA.
AC   P25390;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Serine/threonine protein kinase SSK22 (EC 2.7.-.-).
GN   SSK22 OR YCR073C OR YCR73C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
```

```
RA   Sanz E.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95350642; PubMed=7624781;
RA   Maeda T., Takekawa M., Saito H.;
RT   "Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-
RT   containing osmosensor.";
RL   Science 269:554-558(1995).
CC   -!- FUNCTION: KINASE INVOLVED IN A SIGNAL TRANSDUCTION PATHWAY THAT IS
CC       ACTIVATED BY CHANGES IN THE OSMOLARITY OF THE EXTRACELLULAR
CC       ENVIRONMENT. ACTIVATES THE PBS2 MAP KINASE KINASE BY
CC       PHOSPHORYLATION.
CC   -!- SUBUNIT: INTERACTS WITH BY SSK1.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MAP KINASE KINASE KINASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; CAA42271.1; -.
DR   PIR; S19488; S19488.
DR   HSSP; P24941; 1B38.
DR   SGD; S0000669; SSK22.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN     1034   1314       PROTEIN KINASE.
FT   NP_BIND    1040   1048       ATP (BY SIMILARITY).
FT   BINDING    1063   1063       ATP (BY SIMILARITY).
FT   ACT_SITE   1158   1158       BY SIMILARITY.
SQ   SEQUENCE   1314 AA;  150654 MW;  8040A31FAAD56E14 CRC64;

  Query Match             6.0%;  Score 90.5;  DB 1;  Length 1314;
  Best Local Similarity   19.2%;  Pred. No. 2;
  Matches   61;  Conservative   58;  Mismatches  127;  Indels   71;  Gaps   14;

Qy       3 EQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERM----AIRQTW-GKERMVK 57
           |||  | :| :     :    | |   :| : :    | : |    ::   |   ::| |
Db     209 EQSLTYLRDSSDSVFEEIMKFQIPQGKILSLDALEAILQDLMNRYHSVVSYWPNLKKMYK 268

Qy      58 GKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
            | :         |::|  |  :|         :   :|:  ||||:     :|::|| |
Db     269 DKPI--------TNTAEFTARID---------VMNSWLNFKTNLTLRRQELDDWINRFSP 311

Qy     118 QAAFVMKTDS---------DMFINVDYLTELLLKKNRTTRFFTG----------FLKLN 157
             :  :|:             :  | | |:|:      |              |||: :
Db     312 ----ISSSDNCQEDFDGVPQWNCKMKILAEQLMKEKNIESIFQKKIFYPLSPWMFKLKLH 367

Qy     158 EFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQV-YNVSKSVPYIKLEDVFV 216
             |:  :|  :   :||::|     :   |:    : ::: |     | : : |  :
Db     368 FIVYRETLTKMNI---KYPYERLRSLLAFPVYLIKEVILTRLSYARKLKNPTMMMIDQMI 424

Qy     217 G--LCLERLNIRLE----------ELHSQPTF---FPGGLRFSVCLFR-RIVAC---H 255
               ||:::|:            ::   |||      :|:   |   :::     :
Db     425 DDFNAFIRLSVQLKYTLTKYCSNLPFDVDFDPTFENTVIEAIRYLFFLLNLKLIDSSKQN 484

Qy     256 FIKPRTLLDYWQALENS 272
           |  |  || ||  |:|:
Db     485 FKAPDLLLKYWDHLKNT 501


RESULT 11
```

```
TLR1_HUMAN
ID   TLR1_HUMAN     STANDARD;      PRT;   786 AA.
AC   Q15399; O15452; Q9UG90;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Toll-like receptor 1 precursor (Toll/interleukin-1 receptor-like)
DE   (TIL).
GN   TLR1 OR KIAA0012.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Erythroleukemia;
RX   MEDLINE=98118556; PubMed=9435236;
RA   Rock F.L., Hardiman G., Timans J.C., Kastelein R.A., Bazan J.F.;
RT   "A family of human receptors structurally related to Drosophila
RT   Toll.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:588-593(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051387; PubMed=7584026;
RA   Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA   Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. I.
RT   The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT   analysis of randomly sampled cDNA clones from human immature myeloid
RT   cell line KG-1.";
RL   DNA Res. 1:27-35(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21154917; PubMed=11230166;
RA   Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA   Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA   Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA   Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA   Wambutt R., Korn B., Klein M., Poustka A.;
RT   "Towards a catalog of human genes and proteins: sequencing and
RT   analysis of 500 novel complete protein coding human cDNAs.";
RL   Genome Res. 11:422-435(2001).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF TIR DOMAIN.
RX   MEDLINE=20531768; PubMed=11081518;
RA   Xu Y., Tao X., Shen B., Horng T., Medzhitov R., Manley J.L., Tong L.;
RT   "Structural basis for signal transduction by the Toll/interleukin-1
RT   receptor domains.";
RL   Nature 408:111-115(2000).
CC   -!- FUNCTION: Participates in the innate immune response to microbial
CC       agents. Cooperates with TLR2 and modulates the response to
CC       microbial constituents. Acts via MyD88 and TRAF6, leading to NF-
CC       kappa-B activation, cytokine secretion and the inflammatory
CC       response (By similarity).
CC   -!- SUBUNIT: Binds TLR2 via their respective extracellular domains.
CC       Binds MyD88 via their respective TIR domains (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Plasma membrane and
CC       phagosomes (By similarity).
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in spleen, ovary,
CC       peripheral blood leukocytes, thymus and small intestine.
CC   -!- SIMILARITY: BELONGS TO THE TOLL-LIKE RECEPTOR FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TIR DOMAIN.
CC   -!- SIMILARITY: CONTAINS 8 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; U88540; AAC34137.1; -.
DR   EMBL; D13637; BAA02801.1; -.
DR   EMBL; AL050262; CAB43364.1; -.
DR   Genew; HGNC:11847; TLR1.
DR   MIM; 601194; -.
DR   PDB; 1FYW; 22-NOV-00.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR000157; TIR_domain.
DR   Pfam; PF00560; LRR; 7.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01582; TIR; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00369; LRR_TYP; 4.
DR   SMART; SM00255; TIR; 1.
DR   PROSITE; PS50104; TIR; 1.
KW   Receptor; Immune response; Inflammatory response; Signal;
KW   Transmembrane; Repeat; Leucine-rich repeat; Glycoprotein;
KW   3D-structure.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    786       TOLL-LIKE RECEPTOR 1.
FT   DOMAIN       23    580       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    581    601       POTENTIAL.
FT   DOMAIN      602    786       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       68     91       LRR 1.
FT   REPEAT       93    113       LRR 2.
FT   REPEAT      114    137       LRR 3.
FT   REPEAT      139    164       LRR 4.
FT   REPEAT      371    395       LRR 5.
FT   REPEAT      397    423       LRR 6.
FT   REPEAT      444    467       LRR 7.
FT   REPEAT      469    491       LRR 8.
FT   DOMAIN      635    779       TIR.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    137    137       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    429    429       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     80     80       R -> T (IN REF. 2).
FT   CONFLICT    182    182       E -> G (IN REF. 3).
FT   CONFLICT    228    228       N -> S (IN REF. 3).
FT   CONFLICT    248    248       S -> N (IN REF. 2).
FT   CONFLICT    276    276       F -> S (IN REF. 3).
FT   CONFLICT    602    602       I -> S (IN REF. 2).
SQ   SEQUENCE   786 AA;  90290 MW;  1406C60FB5847F58 CRC64;

  Query Match             6.0%;  Score 89.5;  DB 1;  Length 786;
  Best Local Similarity   22.4%;  Pred. No. 1.3;
  Matches   49;  Conservative   38;  Mismatches   79;  Indels   53;  Gaps   12;

Qy      91 QKDFLDVYYNLTLK--TMMGIEWVHRFCPQ--AAFV-------MKTDSDMFINVDYLTE- 138
           |    ::: |: :|  |: |   ||  ||   : |:       : ||: :| |  :|||
Db     316 QSYIYEIFSNMNIKNFTVSGTRMVHMLCPSKISPFLHLDFSNNLLTDT-VFENCGHLTEL 374

Qy     139 --LLLKKNRTTRFFTGFLKLNEFPIR-QPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDV 195
             |:|: |:         |: |   : :   :  :|::   :|     || |  : | ::
Db     375 ETLILQMNQLKE----LSKIAEMTTQMKSLQQLDISQNSVSYDEKKGDCSWTKSLLSLNM 430

Qy     196 ASQVY------------------NVSKSVP--YIKLEDVFVGLCLERLNIRLEELHSQP 234
           :| :                    |  ||:|   :|||       |: ||:    |
Db     431 SSNILTDTIFRCLPPRIKVLDLHSNKIKSIPKQVVKLE------ALQELNVAFNSL---- 480

Qy     235 TFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALENSR 273
           |  ||   ||     ::  |        |::|: :  |
Db     481 TDLPGCGSFS--SLSVLIIDHNSVSHPSADFFQSCQKMR 517


RESULT 12
```

```
YHV0_YEAST
ID   YHV0_YEAST     STANDARD;      PRT;   579 AA.
AC   P38848;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Hypothetical 66.1 kDa protein in IMP3-SPO12 intergenic region.
GN   YHR150W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U10397; AAB68980.1; -.
DR   PIR; S46757; S46757.
DR   SGD; S0001193; YHR150W.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM    237    253       POTENTIAL.
FT   TRANSMEM    397    421       POTENTIAL.
SQ   SEQUENCE   579 AA;  66148 MW;  2A6A73999C54C19E CRC64;

  Query Match             5.9%;  Score 88.5;  DB 1;  Length 579;
  Best Local Similarity   27.4%;  Pred. No. 1.1;
  Matches   37;  Conservative   22;  Mismatches   51;  Indels   25;  Gaps    7;

Qy      51 GKERMVKGKQL--KTFF--LLGTTSSAAETKEVDQESQRHGDIIQ--KDFLDV--YYNLT 102
           ||     |:|   |:|   |       |  |:|: ::  |:: |  |  |  :::    |:|
Db     304 GKNEYSYGQQFFSKSFMDTLESRNQEIDEISELDKRTENTGELKQGMKVLINLRDMQNMT 363

Qy     103 LKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIR 162
             |:  || ::      :|: |: |  | |         ::  | ||||||| :    |
Db     364 SGTLHVIEAIN------SFLRKSSS--FQN---------EECSTKRFFTGFLLIVFLKIL 406

Qy     163 QPFSKWFVSKSEYPW 177
           || |    | : |
Db     407 SPFVNWSYVCSIFAW 421


RESULT 13
YVNF_AZOCH
ID   YVNF_AZOCH     STANDARD;      PRT;   516 AA.
AC   P24423;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein in vnfD 5'region.
OS   Azotobacter chroococcum mcd 1.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Azotobacter.
OX   NCBI_TaxID=355;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90356423; PubMed=2388847;
RA   Fallik E., Robson R.L.;
RT   "Completed sequence of the region encoding the structural genes for
RT   the vanadium nitrogenase of Azotobacter chroococcum.";
RL   Nucleic Acids Res. 18:4616-4616(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51756; CAA36057.1; -.
DR   PIR; S14694; S14694.
KW   Nitrogen fixation; Hypothetical protein.
SQ   SEQUENCE   516 AA;  58676 MW;  ED7A7F362ED08A14 CRC64;

  Query Match             5.8%;  Score 86.5;  DB 1;  Length 516;
  Best Local Similarity   24.8%;  Pred. No. 1.5;
  Matches   66;  Conservative   30;  Mismatches   97;  Indels   73;  Gaps   12;

Qy      25 TPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQ--- 81
           || ||| |:|       || |   |   |:::    | : | :|  |: |   :
Db     284 TPTFLVDLLT-------ERRAF--TPALEQVMATDALLSAFEVGDISTEALMFQAGYLTI 334

Qy      82 -ESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 140
            : || |   |  :   | || :|  :    : |: |: | ||:    ::       :||
Db     335 GQVQRIG--AQPRYRLRYPNLEVKASLNEALLLRYLPEPASVMRQTGQLY-------DLL 385

Qy     141 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 200
           |               |:||  |     | :   : | |  |     ||      ||  |
Db     386 L--------------ANDFPGLQTLFTAFFASIPHDWYRNNPIARYEGY-----YASVFY 426

Qy     201 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 260
           :             |  |    |::|||:  |      | :  :| ||   |      |
Db     427 S------------HFAAL---GLDVRLEDATSH-----GRIDMTV-LFNGHVYLFEFKVV 465

Qy     261 TLLDYWQALE-----------NSRGE 275
             |    |||:           :|||
Db     466 ELTPAGQALQQLKDKGYADKYRARGE 491


RESULT 14
HSER_PIG
ID   HSER_PIG       STANDARD;      PRT;  1073 AA.
AC   P55204; Q29050;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Heat-stable enterotoxin receptor precursor (GC-C) (Intestinal
DE   guanylate cyclase) (EC 4.6.1.2) (STA receptor).
GN   GUCY2C OR GUC2C.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Small intestine;
RX   MEDLINE=95058376; PubMed=7968686;
RA   Wada A., Hirayama T., Kitao S., Fujisawa J.-I., Hidaka Y.,
RA   Shimonishi Y.;
RT   "Pig intestinal membrane-bound receptor (guanylyl cyclase) for heat-
RT   stable enterotoxin: cDNA cloning, functional expression, and
RT   characterization.";
RL   Microbiol. Immunol. 38:535-541(1994).
RN   [2]
RP   REVISIONS TO 238 AND 509.
RA   Wada A.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR THE E.COLI HEAT-STABLE ENTEROTOXIN (E.COLI
CC       ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN
CC       MAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
CC       PEPTIDE GUANYLIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: GTP = 3',5'-cyclic GMP + diphosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17513; BAA04465.1; -.
DR   HSSP; Q02846; 1AWL.
DR   InterPro; IPR001828; ANF_receptor.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001054; G_cyclase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00211; guanylate_cyc; 1.
DR   Pfam; PF01094; ANF_receptor; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Receptor; Transmembrane; Glycoprotein; Lyase; cGMP synthesis; Signal.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24   1073       HEAT-STABLE ENTEROTOXIN RECEPTOR.
FT   DOMAIN      24    433       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   434    454       POTENTIAL.
FT   DOMAIN     455   1073       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     489    749       PROTEIN KINASE-LIKE.
FT   DOMAIN     824    954       GUANYLATE CYCLASE.
FT   CARBOHYD    32     32       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   284    284       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   307    307       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   357    357       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1073 AA;  123219 MW;  E6F4FC1327BA9F43 CRC64;

  Query Match             5.7%;  Score 85;  DB 1;  Length 1073;
  Best Local Similarity   23.1%;  Pred. No. 5.2;
  Matches   27;  Conservative   15;  Mismatches   43;  Indels   32;  Gaps    3;

Qy      95 LDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDM----------FINVDY---LTELLL 141
           ||:   :: :  ||    :    |  : | |  |:|:           :: ||   || |:
Db     103 LDLLRTISSEKRMGCVLLGPSCTYSTFQMYLDTDLNYPMISAGSFGLSCDYKETLTRLMS 162

Qy     142 KKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQ 198
               :   |   | |:| ||                    : ||   | |||     |:
Db     163 PARKLMYFLVDFWKVNNFP------------------FKPFSWNTAYVFKNSTESE 200


RESULT 15
ETF2_MYXVL
ID   ETF2_MYXVL     STANDARD;      PRT;   711 AA.
AC   Q9Q8K4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Early transcription factor 82 kDa subunit (VETF large subunit).
GN   M096L.
OS   Myxoma virus (strain Lausanne).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
```

```
OC   Leporipoxvirus.
OX   NCBI_TaxID=31530;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20032073; PubMed=10562494;
RA   Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.X.,
RA   Macaulay C., Willer D., Evans D., McFadden G.;
RT   "The complete DNA sequence of myxoma virus.";
RL   Virology 264:298-318(1999).
CC   -!- FUNCTION: ACTS WITH RNA POLYMERASE TO INITIATE TRANSCRIPTION FROM
CC       EARLY GENE PROMOTERS. A DNA-DEPENDENT ATPASE ACTIVITY IS
CC       ASSOCIATED WITH VETF.
CC   -!- SUBUNIT: HETERODIMER OF A 70 kDa AND A 82 kDa SUBUNIT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF170726; AAF14984.1; -.
KW   Transcription regulation; Activator.
SQ   SEQUENCE   711 AA;  82687 MW;  F69EC4F63B087E61 CRC64;

  Query Match             5.6%;  Score 84.5;  DB 1;  Length 711;
  Best Local Similarity   20.1%;  Pred. No. 3.5;
  Matches   64;  Conservative   54;  Mismatches  115;  Indels   85;  Gaps   16;

Qy       4 QSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKT 63
           :| :|| |  |: |       |   | | |  |     : :    : |   |:   :| |
Db     206 KSLLYKSDEYFINLTGNKFILTDERLNLTVWDS-----DGVVTFSSDGDTIMINNVKLFT 260

Qy      64 FFLLGTTSSAAETKEVDQESQR-HGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFV 122
             |           ::| : :|  ||:  | ||     :| :  : ||          ||
Db     261 SLL----------TDIDLQMERIKGDVTYKIFLST--PITSRIKLNIE------TSFIFV 302

Qy     123 MKTDSDMFINVDYLTELLLKKNRTT----------RFFTGFLKL---NEFPIRQPFSKW 168
              ::: :: |    ::| ||  :           ::|| || :   : |   |  | :
Db     303 ETATNNILLSADKRISIILAKNHISIKVKNYIPNIEKYFT-FLVIAINSMFNNIQQASDF 361

Qy     169 FVSKSEYPWDRYPPFCSGTGY-----VFSGDVASQVYNVSKSVPYIKLEDVFVG------ 217
             :: | | |    |  |       |    :   :  || :      ::||:
Db     362 TKVETVY-WSR---ICQNTKNKHRKPVIVSSLDEDMEKVSDNFYKSSTKEVFINSSGIMF 417

Qy     218 LCLERLN----------IRLEEL----------HSQPTFFPGGLRFSVCLFRRIVACH 255
             ||: ||           ||:::          |::        || |:::: :|
Db     418 SCLDPLNKYNYVGFLSIFYRLQKMCIPCCFLKNQSHTE--------TFSSCVYQKEIASD 469

Qy     256 FIKPRTLLDYWQALENSR 273
           | |  :|:: : :  |:
Db     470 VINP-YILNFGKVVTKSK 486


Search completed: April  4, 2003, 14:44:27
Job time : 14 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:42:35 ; Search time 19 Seconds
(without alignments)
1416.718 Million cell updates/sec

Title: US-09-914-152-1_COPY_31_310
Perfect score: 1498
Sequence: 1 FKEQSFVYKKDGNFLKLPDT.....TLLDYWQALENSRGEDCPPV 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	291.5	19.5	684	2	F86394	protein T24P13.20
2	267	17.8	657	2	G86397	protein T7N9.18 [i
3	264.5	17.7	739	2	T10648	hypothetical prote
4	243.5	16.3	642	2	D96777	hypothetical prote
5	233.5	15.6	262	2	T20203	hypothetical prote
6	223.5	14.9	384	2	T20446	hypothetical prote
7	219.5	14.7	507	2	T18639	hypothetical prote
8	200	13.4	357	2	T20029	hypothetical prote
9	199.5	13.3	345	2	T20031	hypothetical prote
10	199	13.3	376	2	T24925	hypothetical prote
11	192	12.8	332	2	T20028	hypothetical prote
12	189.5	12.7	376	2	T20876	hypothetical prote
13	171.5	11.4	398	2	B86353	protein F2E2.6 [im
14	160	10.7	353	2	C96573	protein F12M16.19
15	158.5	10.6	253	2	T22387	hypothetical prote
16	155.5	10.4	325	2	T24762	hypothetical prote
17	146.5	9.8	409	2	A84733	hypothetical prote
18	143.5	9.6	404	2	A86186	hypothetical prote
19	140.5	9.4	338	2	T24743	hypothetical prote
20	139.5	9.3	399	2	H86453	avr9 homolog F9L11
21	139	9.3	406	2	T04817	hypothetical prote
22	137.5	9.2	359	2	F87937	protein F14B6.6 [i
23	137.5	9.2	385	2	T20879	hypothetical prote
24	134	8.9	395	2	A86458	probasble elicitor
25	128.5	8.6	348	2	T31918	hypothetical prote
26	122.5	8.2	401	2	A86251	hypothetical prote
27	115	7.7	334	2	T32256	hypothetical prote
28	109.5	7.3	333	2	T02614	hypothetical prote
29	105.5	7.0	753	2	T24745	hypothetical prote
30	105.5	7.0	1049	2	S67613	vacuolar carboxype
31	104	6.9	414	2	B96808	protein F28K19.2 [
32	101	6.7	349	2	T24744	hypothetical prote
33	99	6.6	845	2	T46164	nodulin / glutamat
34	98	6.5	660	2	T22794	hypothetical prote
35	93.5	6.2	336	2	S70708	site-specific DNA-
36	93	6.2	368	2	T15096	hypothetical prote
37	93	6.2	3724	2	T18427	hypothetical prote
38	92.5	6.2	378	2	T39790	hypothetical prote
39	91	6.1	279	2	T20443	hypothetical prote
40	90.5	6.0	338	2	T49998	hypothetical prote
41	90.5	6.0	1314	2	S19488	probable membrane
42	90	6.0	482	2	F82828	phosphomannose iso
43	90	6.0	786	2	T08664	Toll protein-like
44	88.5	5.9	579	2	S46757	hypothetical prote
45	86.5	5.8	516	2	S14694	hypothetical prote

ALIGNMENTS

RESULT 1
F86394
protein T24P13.20 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: F86394
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Al
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Ki
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marz
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tal
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86394
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-684 <STO>
A;Cross-references: GB:AE005172; NID:g9295733; PIDN:AAF87039.1; GSPDB:GN00141
C;Genetics:
A;Gene: T24P13.20
A;Map position: 1

Query Match 19.5%; Score 291.5; DB 2; Length 684;
Best Local Similarity 28.7%; Pred. No. 7.2e-19;
Matches 79; Conservative 49; Mismatches 120; Indels 27; Gaps 6;

```
Qy      26 PPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQR 85
           |  ||: | |:       |||:|:|| :   |:  ::   | :|   |     |:  |::
Db     410 PLDLVIGVFSTANNFKRRMAVRRTWMQYDDVRSGRVAVRFFVGLHKSPLVNLELWNEART 469

Qy      86 HGDIIQKDFLDVYYNLTLKTM----MGIEWVHRFC-----------PQAAFVMKTDSDMF 130
           :||:    |:| |  :: ||:     |:  :: |             | |:|||| | |
Db     470 YGDVQLMPFVDYYSLISWKTLAICIFGLSSLYCFASLTSYFKQTEVDSAKFIMKTDDDAF 529

Qy     131 INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVS------KSEYPWDRYPPFC 184
           : || :   |   | |     | :  :  ||| | |||::|      | |:| ::|||:
Db     530 VRVDEVLLSLSMTNNTRGLIYGLINSDSQPIRNPDSKWYISYEFLLKKQEWPEEKYPPWA 589

Qy     185 SGTGYVFSGDVASQVYNVSK--SVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLR 242
           | ||: | |:|  |  : |  ::   |||||  :|: :  |     | | :     | :
Db     590 HGPGYIVSRDIAESVGKLFKEGNLKMFKLEDVAMGIWIAELTKHGLEPHYEN---DGRII 646

Qy     243 FSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
           |    :|| |:  |  :   |:  : ::   |
Db     647 SDGCKDGYVVA-HYQSPAEMTCLWRKYQETKRSLC 680
```

RESULT 2

```
G86397
protein T7N9.18 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86397
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86397
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-657 <STO>
A;Cross-references: GB:AE005172; NID:g8778858; PIDN:AAF79857.1; GSPDB:GN00141
C;Genetics:
A;Gene: T7N9.18
A;Map position: 1

  Query Match             17.8%;  Score 267;  DB 2;  Length 657;
  Best Local Similarity   27.5%;  Pred. No. 1.3e-16;
  Matches   67;  Conservative   51;  Mismatches  102;  Indels   24;  Gaps    7;

Qy      43 RMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLT 102
           |||:|::| ::::|:  ::   | :   :       :: :|::  |||:   ::| |  :
Db     425 RMAVRKSWMQQKLVRSSKVVARFFVALHARKEVNVDLKKEAEYFGDIVIVPYMDHYDLVV 484

Qy     103 LKTM----MGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNE 158
           |||:     |:  |      | :||| | | |: || : :   |       : | :   |
Db     485 LKTVAICEYGVNTV-----AAKYVMKCDDDTFVRVDAVIQEAEKVKGRESLYIGNINFNH 539

Qy     159 FPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN--VSKSVPYIKLEDVFV 216
            |:|    || |:  |:| : |||: :| ||: | |||  : :   | :   |:||| :
Db     540 KPLRT--GKWAVTFEEWPEEYYPPYANGPGYILSYDVAKFIVDDFEQKRLRLFKMEDVSM 597

Qy     217 GLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRI---VACHFIKPRTLLDYWQALENSR 273
           |: :|: |       ::|      |:|  | |  |      |:  || ::  |  |:
Db     598 GMWVEKFN------ETRPVAVVHSLKF--CQFGCIEDYFTAHYQSPRQMICMWDKLQRLG 649

Qy     274 GEDC 277
           |
Db     650 KPQC 653


RESULT 3
T10648
hypothetical protein T13K14.220 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C;Accession: T10648
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10648
A;Molecule type: DNA
A;Residues: 1-739 <BEV>
A;Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.220
A;Experimental source: cultivar Columbia; BAC clone T13K14
C;Genetics:
A;Gene: ATSP:T13K14.220
A;Map position: 4
A;Introns: 59/3; 86/1; 358/1; 438/3; 535/3; 578/3; 634/3; 670/3

  Query Match             17.7%;  Score 264.5;  DB 2;  Length 739;
  Best Local Similarity   27.6%;  Pred. No. 2.6e-16;
  Matches   71;  Conservative   48;  Mismatches  123;  Indels   15;  Gaps    6;

Qy      25 TPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQ 84
           ||  | : | |:    :||||:|:|| :   :|   :   | :           : :|::
Db     491 TPFRLFMGVLSATNHFSERMAVRKTWMQHPSIKSSDVVARFFVALNPRKEVNAMLKKEAE 550

Qy      85 RHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKN 144
           |||:   |:| |  : |||:   |: :   |   ::|| | | || |: : : :   :
Db     551 YFGDIVILPFMDRYELVVLKTIAICEFGNVTAP---YIMKCDDDTFIRVESILKQIDGVS 607

Qy     145 RTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQV--YNV 202
               : | | |   |:|    || |:  |:|   |||: :| ||: | ::|   :   |
Db     608 PEKSLYMGNLNLRHRPLRT--GKWTVTWEEWPEAVYPPYANGPGYIISSNIAKYIVSQNS 665

Qy     203 SKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSV--CLFRRIVACHFIKPR 260
              :  |:||| :|| :|: |  :     ||  :    :|    |      | |:  |
Db     666 RHKLRLFKMEDVSMGLWVEQFNASM-----QPVEYSHSWKFCQYGCTLNYYTA-HYQSPS 719

Qy     261 TLLDYWQALENSRGEDC 277
           ::  |  |   | : |
Db     720 QMMCLWDNLLKGRPQCC 736


RESULT 4
D96777
hypothetical protein F25A4.23 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96777
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alc
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kin
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-642 <STO>
A;Cross-references: GB:AE005173; NID:g5882743; PIDN:AAD55296.1; GSPDB:GN00141
C;Genetics:
A;Gene: F25A4.23
A;Map position: 1

  Query Match             16.3%;  Score 243.5;  DB 2;  Length 642;
  Best Local Similarity   24.9%;  Pred. No. 1.9e-14;
  Matches   65;  Conservative   51;  Mismatches  116;  Indels   29;  Gaps    8;

Qy      28 FLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHG 87
           |: :|   :|   :||||:|::| :  ::   ::   | :           |: :|::  |
Db     396 FIGILSAGNH--FSERMAVRKSWMQHVLITSAKVVARFFVALHGRKEVNVELKKEAEYFG 453

Qy      88 DIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA------FVMKTDSDMFINVDYLTELLL 141
           ||:   ::| |  : |||:          |   |       ::|| | | |: :   :   :
Db     454 DIVLVPYMDSYDLVVLKTV-------AICEHGALAFSAKYIMKCDDDTFVKLGAVINEVK 506

Qy     142 KKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN 201
           |       : | :     |:|    || |:  |:| : |||: :| ||| | |:|  : :
Db     507 KVPEGRSLYIGNMNYYHKPLRG--GKWAVTYEEWPEEDYPPYANGPGYVLSSDIARFIVD 564

Qy     202 VSK--SVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLF---RRIVACHF 256
             :   :  |:||| ||: :|       :  : |  :   |||  | |          |:
Db     565 KFERHKLRLFKMEDVSVGMWVEHF-----KNTTNPVDYRHSLRF--CQFGCVENYYTAHY 617

Qy     257 IKPRTLLDYWQALENSRGEDC 277
           || ::  |  |       :|
Db     618 QSPRQMICLWDKLLRQNKPEC 638


RESULT 5
```

```
T20203
hypothetical protein C54C8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20203
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19237
A;Accession: T20203
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-262 <WIL>
A;Cross-references: EMBL:Z83102; PIDN:CAB05465.1; GSPDB:GN00019; CESP:C54C8.3
A;Experimental source: clone C54C8
C;Genetics:
A;Gene: CESP:C54C8.3
A;Map position: 1
A;Introns: 19/2; 117/3; 184/2; 226/3

  Query Match             15.6%;  Score 233.5;  DB 2;  Length 262;
  Best Local Similarity   27.7%;  Pred. No. 5e-14;
  Matches   67;  Conservative   46;  Mismatches  100;  Indels   29;  Gaps    6;

Qy      13 NFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFFLLGT 69
           ::| ||: :       ::::| |     | |  :||||       :|   ::|  ||:|
Db      22 SWLYLPEIEETSQEKDILIIVASRTDSYARRNILRQTWMSKANSEIVANGRMKPLFLVGL 81

Qy      70 TSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSD 128
           |    : |: | ||:: :|||| |  | | |  || |::  : :     |:   : | | |
Db      82 TPGEYKMKKMVMQEAKLYGDIIVVDMNDNYEELTYKSLAILLYGVSKAPRYQMIGKIDED 141

Qy     129 MFINVDYLTELLLKKNRTTRFFTGFLKLNEFPI------------RQPFSKWFVSKSEYP 176
           :    | ||||          :  ||:       |             |    :|:| :| |
Db     142 VMFFPDKLTEL---------YDQGFIDATPLRIYGLKMQSGANIFRDKTHRWYVPESSYS 192

Qy     177 WDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEELHSQPT 235
             ::| : ||  |: : : | |:   :|   :|::||||: |:  | | | :: |   |
Db     193 CSKFPEYVSGMLYMVTWEAAQQIIKSTKYRDFIQVEDVFLTGILAEDLGISVKNL---PE 249

Qy     236 FF 237
           |:
Db     250 FY 251


RESULT 6
T20446
hypothetical protein E03H4.11 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T20446
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19276
A;Accession: T20446
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-384 <WIL>
A;Cross-references: EMBL:Z81492; PIDN:CAB04032.1; GSPDB:GN00019; CESP:E03H4.11
A;Experimental source: clone E03H4
C;Genetics:
A;Gene: CESP:E03H4.11
A;Map position: 1
A;Introns: 46/1; 76/1; 196/3; 263/2; 305/3; 335/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             14.9%;  Score 223.5;  DB 2;  Length 384;
  Best Local Similarity   27.0%;  Pred. No. 7.1e-13;
  Matches   67;  Conservative   44;  Mismatches  104;  Indels   33;  Gaps    6;

Qy       9 KKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFF 65
           :|:  :| ||: :       ::::| |     | |  :||||       :|   ::|  |
Db      97 QKNYTWLYLPEIEEASQEKDILMIVASRTDSYARRNIMRQTWMNKSDSEIVANGRMKPLF 156

Qy      66 LLGTTSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMK 124
           |:| |    : |: | ||:: :|||| |  | | |  || |::  : :     |:   : |
Db     157 LVGLTPGDYKMKKMVMQEAKLYGDIIVVDMNDTYEELTYKSLAILLYGVSKAPRYQMIGK 216

Qy     125 TDSDMFINVDYLTELL-------------LKKNRTTRFFTGFLKLNEFPIRQPFSKWFV 170
           | |:   | || |                |     | |           |   :|:|
Db     217 IDEDVIFFPDKLTALYEQGIIDATPLCAYGYKIQAGARIF----------RDKNDRWYV 265

Qy     171 SKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEE 229
           :| |   ::| : ||  |: : : | |:   :|   :|::|||: |:  | | | :
Db     266 PESSYSCSKFPEYVSGMLYMVTWEAAQQIIKSTKYRDFIQVEDVFLTGILAEDLGISVRN 325

Qy     230 LHSQPTFF 237
           |   | |:
Db     326 L---PKFY 330


RESULT 7
T18639
hypothetical protein B0024.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18639
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19001
A;Accession: T18639
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-507 <WIL>
A;Cross-references: EMBL:Z71178; PIDN:CAA94876.1; GSPDB:GN00023; CESP:B0024.3
A;Experimental source: clone B0024
C;Genetics:
A;Gene: CESP:B0024.3
A;Map position: 5
A;Introns: 69/2; 118/2; 143/2; 181/1; 246/2; 281/3; 393/2; 437/3

  Query Match             14.7%;  Score 219.5;  DB 2;  Length 507;
  Best Local Similarity   25.6%;  Pred. No. 2.4e-12;
  Matches   74;  Conservative   56;  Mismatches  138;  Indels   21;  Gaps    9;

Qy       1 FKEQSFVYKKDGNFLKLPDTDC-RQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMV 56
           ||:|:  |:    |: :|   |   |   : :| |: :  :  | |||:||       |
Db     227 FKDQNHTYQ----FITVPKKQCSNNTKLQITILSTAGNFDI--RQAIRETWANPNNSEHV 280

Qy      57 KGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFC 116
              ::  |::  ||:       : :| ::  |:|  |  : |  | ||    : :    |
Db     281 ANNDVRISFIISKTSNEFLNFALQKEIEKFDDMIVTDLYESYELLILKVHAILSYKQSHC 340

Qy     117 PQAAFVMKTDSDMFINVDYLTELLL-KKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEY 175
             | | :| | || :::| |   |  ||  :   :| :  |  |:|:   :|:| |: |
Db     341 QLADFQLKIDDDMAVDMDGLYRSLEDKKQASINGISGIIWKNSPPVREKKHRWYVPKTLY 400

Qy     176 PWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVF----VG--LCLERLNIRLEE 229
           :||:  |  |:   :   ::   :|:     :||||    :|  | ::::|
Db     401 SEKFFPPYIDGPIYLIGKNAVPRMLEEAKNYNQWIIEDVFWTGVIGKALKIKQINWANHL 460

Qy     230 LHSQPTFFPGGLRFSVCLFRRIVACHFIK-PRTLLDYWQALENSRGEDC 277
           |       |  |: |       | | | :| |: : | :| |   :|  |
Db     461 LRYVIELIPSRLKCSKGGVPLIYAVHNMKGPQMIHDGYQKL---KGVKC 506


RESULT 8
T20029
hypothetical protein C47F8.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T20029
R;McLay, K.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z19212
A;Accession: T20029
A;Status: preliminary; translated from GB/EMBL/DDBJ
```

```
A;Molecule type: DNA
A;Residues: 1-357 <WIL>
A;Cross-references: EMBL:AL009246; PIDN:CAA15839.1; GSPDB:GN00019; CESP:C47F8.6
A;Experimental source: clone C47F8
C;Genetics:
A;Gene: CESP:C47F8.6
A;Map position: 1
A;Introns: 19/1; 52/1; 172/3; 239/2; 281/3; 311/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             13.4%;  Score 200;  DB 2;  Length 357;
  Best Local Similarity   25.7%;  Pred. No. 9.8e-11;
  Matches   59;  Conservative   48;  Mismatches  115;  Indels    8;  Gaps    4;

Qy       9 KKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFF 65
Db      73 QKNYTWLYLPTIEETSKETDILMIVASRTDSYARRNIMRQTWMNKSNSEIVANGRMKSLF 132

Qy      66 LLGTTSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMK 124
Db     133 LVGLAPADYKVKKMVMQEAKLYGDIIIVDMDDTYEELIYKSLMIFLFGVSKAPQYKIIGK 192

Qy     125 TDSDMFINVDYLTELL---LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYP 181
Db     193 IDEDIMFFPDKLMALYEQGIIDSTPVSLYGLVIPAGRDIFRDKTNRWYVPESAYSCSQYP 252

Qy     182 PFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEEL 230
Db     253 AYLSGMYYMATREAAQMLLKSTKHRDFIQVEDVLLTGILAEDLGIPIRDM 302


RESULT 9
T20031
hypothetical protein C47F8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T20031; T22806
R;McLay, K.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z19212
A;Accession: T20031
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-345 <WIL>
A;Cross-references: EMBL:AL009246; PIDN:CAA15841.1; GSPDB:GN00019; CESP:C47F8.3
A;Experimental source: clone C47F8
R;Kershaw, J.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19617
A;Accession: T22806
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-345 <WI2>
A;Cross-references: EMBL:Z81553; PIDN:CAB04502.1; GSPDB:GN00019; CESP:C47F8.3
A;Experimental source: clone F56H6
C;Genetics:
A;Gene: CESP:C47F8.3
A;Map position: 1
A;Introns: 23/1; 49/1; 169/3; 236/2; 278/3; 308/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             13.3%;  Score 199.5;  DB 2;  Length 345;
  Best Local Similarity   27.6%;  Pred. No. 1e-10;
  Matches   60;  Conservative   39;  Mismatches  107;  Indels   11;  Gaps    7;

Qy      29 LVLLVTSSHKQLAERMAIRQTW-GKE--RMVKGKQLKTFFLLGTTSSAAETKE-VDQESQ 84
Db      90 ILMLVVSKTKNFARRNVLRSTWMNKENSEMMKSGRMHALFFVGLVPGDQNLKKLVLEEAE 149

Qy      85 RHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKK- 143
Db     150 IHGDMVVVDLEDTYDNLPFKTLALLLYGTSKASQFKIIGKIDDDVMFFPDQLLPMLDRNF 209

Qy     144 -NRTTRFFTGFLK-LNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN 201
Db     210 VNSNTLSIYGHLSTAEELVLRNKTEPWYVPETAYNCTVYPVYVMGPIYLVTKDAASLILD 269

Qy     202 VSKSVPYIKLEDVFV-GLCLERLNIRLEELHSQPTFF 237
Db     270 NANHQQFMTVEDALITGIIAQKLGIR---RYSLPNVF 303


RESULT 10
T24925
hypothetical protein T15D6.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T24925
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19956
A;Accession: T24925
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-376 <WIL>
A;Cross-references: EMBL:Z83125; PIDN:CAB05616.1; GSPDB:GN00019; CESP:T15D6.5
A;Experimental source: clone T15D6
C;Genetics:
A;Gene: CESP:T15D6.5
A;Map position: 1
A;Introns: 59/1; 87/1; 207/3; 274/2; 316/3; 339/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             13.3%;  Score 199;  DB 2;  Length 376;
  Best Local Similarity   23.5%;  Pred. No. 1.3e-10;
  Matches   64;  Conservative   60;  Mismatches  120;  Indels   28;  Gaps   10;

Qy       9 KKDGNFLKLPDTDCRQTPPF---LVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLK 62
Db     108 QKSQRWLHLPKI---KNPTWNRDILMIVASRPGSVSRRKVLRKTWMNKANSKIIRNGRMQ 164

Qy      63 TFFLLGTTSSAAE-TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAF 121
Db     165 VLFLVGMVAGDRDLMKAVKKEAESFGDIIVMNLEDTYDNLPFKVLSLLLYGTNKASDFKI 224

Qy     122 VMKTDSDMFINVDYLTELLLKK--NRTTRFFTGFLKL-NEFPIRQPFSKWFVSKSEYPWD 178
Db     225 IGKIDDDVIFFPDRLTPLLDENVIDSSSYSIYGYLSQDDELVVRNETKPWYVPETAYNCT 284

Qy     179 RYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEELHSQPTFF 237
Db     285 KYPVYALGPFYLITNKAANLIVENSRFQNFMTVEDALIAGIIAEGLGI---QRHSLPMNT 341

Qy     238 PGGLRFSVCLFRRIVACHFIK--PRTLLDYWQ 267
Db     342 DG---------KKILSWHMSKRSDRQFLDFYQ 364


RESULT 11
T20028
hypothetical protein C47F8.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T20028
R;McLay, K.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z19212
A;Accession: T20028
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-332 <WIL>
A;Cross-references: EMBL:AL009246; PIDN:CAA15838.1; GSPDB:GN00019; CESP:C47F8.5
A;Experimental source: clone C47F8
C;Genetics:
A;Gene: CESP:C47F8.5
A;Map position: 1
```

```
A;Introns: 26/1; 146/3; 213/2; 255/3; 285/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             12.8%;  Score 192;  DB 2;  Length 332;
  Best Local Similarity   25.7%;  Pred. No. 4.9e-10;
  Matches   59;  Conservative   45;  Mismatches  118;  Indels    8;  Gaps    4;

Qy       9 KKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFF 65
Db      47 QKTYKWLYLPETDIVLRSPDILMMVASRTDSFARRNVLRKTWMNKNYSEIVRDGRMKALF 106

Qy      66 LLGTTSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMK 124
Db     107 LVGMVSEDYRVRRIVMEEAKLYGDMVVIDLEDTYDDLPFKSLSLLLYAVSKAPEFKVIGK 166

Qy     125 TDSDMFINVDYLTELLLKK---NRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYP 181
Db     167 IDEDVMFFPDKLIPLLDGKVIDPDAAAFYGQLLKEGEPVIKKKDAHWYVPDYAYNCTGYP 226

Qy     182 PFCSGTGYVFSGDVASQVYNVSKSVPYIKLED-VFVGLCLERLNIRLEEL 230
Db     227 AYVAGPFYLATRKAAKLVLKFTKFQNFMTVEDSLITGILANDLGIPRKNL 276


RESULT 12
T20876
hypothetical protein F14B6.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T20876
R;White, S.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19338
A;Accession: T20876
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-376 <WIL>
A;Cross-references: EMBL:Z81502; PIDN:CAB04107.1; GSPDB:GN00019; CESP:F14B6.4
A;Experimental source: clone F14B6
C;Genetics:
A;Gene: CESP:F14B6.4
A;Map position: 1
A;Introns: 31/1; 69/1; 189/3; 256/2; 298/3; 328/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             12.7%;  Score 189.5;  DB 2;  Length 376;
  Best Local Similarity   23.6%;  Pred. No. 9.9e-10;
  Matches   68;  Conservative   57;  Mismatches  138;  Indels   25;  Gaps   10;

Qy       9 KKDGNFLKLPDTDCRQT------PPF-------LVLLVTSSHKQLAERMAIRQTW---GK 52
Db      76 KDFGSSFTISFADIQKTHSWLYLPKFQNSMSSEILMIVASRTDSFARRNVLRKTWMNPEN 135

Qy      53 ERMVKGKQLKTFFLLGTT--SSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIE 110
Db     136 SEIIKDGRMKALFLVGMTDGDDSRMRKVVMEEARIYGDMVVVDLKDTYEELPFKSLTTLL 195

Qy     111 WVHRFCPQAAFVMKTDSDMFINVDYLTELLLKK--NRTTRFFTGFL-KLNEFPIRQPFSK 167
Db     196 YGTSKASEFKLIGKIDEDIMFFPDKILPLLEQNLIDPSSESIYGMLFAEGGYVYRDKEHR 255

Qy     168 WFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRL 227
Db     256 WFVPDSTYGCDMFPPYTGGLFYLVTQDAAKKILNATKHRIFIPIEDALINGILAN-DCKI 314

Qy     228 EELHSQPTFFPGGLRFSVCLFRRIVACHFIK-PRTLLDY-WQALENSR 273
Db     315 PRIH-LPEIYGDHHSLTDEDTRGVLAWHTGKNDQQFLDYFYRKLSESR 361


RESULT 13
B86353
protein F2E2.6 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: B86353
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Al
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Ki
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marz
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tal
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B86353
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <STO>
A;Cross-references: GB:AE005172; NID:g9280694; PIDN:AAF86563.1; GSPDB:GN00141
C;Genetics:
A;Gene: F2E2.6
A;Map position: 1

  Query Match             11.4%;  Score 171.5;  DB 2;  Length 398;
  Best Local Similarity   26.0%;  Pred. No. 5e-08;
  Matches   59;  Conservative   45;  Mismatches   80;  Indels   43;  Gaps   11;

Qy      28 FLVLLVTSSHKQLAERMAIRQTW---GK--ERMVKGKQLKTFFLLG--TTSSAAETKEVD 80
Db     127 FMVIGINTAFSSRKRRDSLRETWMPQGEKLEKLEKEKGIVVKFMIGHSSTPNSMLDKEID 186

Qy      81 QESQRHGDIIQKDFLDVYYNLTLKTMMGI-----EWVHRFCPQAAFVMKTDSDMFINVDY 135
Db     187 SEDAQYNDFFRLDHVEGYYNLSAKTKSFFSSAVAKW------DAEFYVKIDDDVHVNLGT 240

Qy     136 LTELLLKKNRTTRFFTGFLKLNEFPI--------RQP-FSKWFVSKSEYPWDRYPPFCSG 186
Db     241 LASTLASHRSKPRVYIGCMKSG--PVLTKKTAKYREPEFWKFGEEGNKY-------FRHA 291

Qy     187 TG--YVFSGDVASQVYNVSKSVPYIKLEDV-----FVGLCLERLNIR 226
Db     292 TGQIYAISKDLATYISNNQPILHKYANEDVTLGSWFIGLEVEQIDDR 338


RESULT 14
C96573
protein F12M16.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96573
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Al
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Ki
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marz
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tal
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C96573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-353 <STO>
A;Cross-references: GB:AE005173; NID:g7769857; PIDN:AAF69535.1; GSPDB:GN00141
C;Genetics:
A;Gene: F12M16.19
A;Map position: 1

  Query Match             10.7%;  Score 160;  DB 2;  Length 353;
  Best Local Similarity   23.6%;  Pred. No. 5e-07;
  Matches   61;  Conservative   42;  Mismatches  112;  Indels   44;  Gaps    8;
```

```
Qy      15 LKLPDTDC-RQTPPFLVLLV--------------TSSHKQL------------AERMAIR 47
Db      47 LRLTGRNCLTNTPPKTVRVVWDVAGNSNGVVSGEKKRHKVMGFVGIQTGFGSAGRRRSLR 106

Qy      48 QTW------GKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNL 101
Db     107 KTWMPSDPEGLRRLEESTGLAIRFMIGKTKSEEKMAQLRREIAEYDDFVLLDIEEEYSKL 166

Qy     102 TLKTMMGIEWVHRFCPQAA-----FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKL 156
Db     167 PYKTLVRVICLAFFKAAYALYDSEFYVKADDDIYLRPDRLSLLLAKERSHSQTYLGCLK- 225

Qy     157 NEFPI-RQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY---NVSKSVPYIKLE 212
Db     226 -KGPVFTDPKLKWYEPLSHLLGKEYFLHAYGPIYALSADVVASLVALKNNRHALLMFNNE 284

Qy     213 DVFVGLCLERLNIRLEELH 231
Db     285 DVTIGAWMLAMNVNHENHH 303


RESULT 15
T22387
hypothetical protein F48F7.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T22387
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19559
A;Accession: T22387
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-253 <WIL>
A;Cross-references: EMBL:Z69661; PIDN:CAA93493.1; GSPDB:GN00028; CESP:F48F7.3
A;Experimental source: clone F48F7
C;Genetics:
A;Gene: CESP:F48F7.3
A;Map position: X
A;Introns: 13/3; 55/3; 93/2; 161/3; 219/2

  Query Match            10.6%;  Score 158.5;  DB 2;  Length 253;
  Best Local Similarity   23.2%;  Pred. No. 4.5e-07;
  Matches   48;  Conservative   45;  Mismatches   87;  Indels   27;  Gaps    5;

Qy      64 FFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVM 123
Db      19 YFIVGTDIDV----DIDDEIKTYEDIIQVDINENYHNITYKAIFWINEINKCRHGPKLLL 74

Qy     124 KTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 183
Db      75 KIDDDVHIDMIGLQFLIKRYSVMDDFMACRVISNGQVVRNNTSKWYLSKKDYKFPNLGTY 134

Qy     184 CSGTGYVFSGDVASQVY-NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLR 242
Db     135 CQGMVYFISGNLLPILHGNIEKS-QFLWMDDWYVTRSL-----------------VGDYK 176

Qy     243 FSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     177 ISYYSLEQ----HSISPNTVREVYSSL 199


Search completed: April  4, 2003, 14:45:30
Job time : 21 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 12:08:21 ; Search time 75 Seconds
(without alignments)
550.769 Million cell updates/sec

Title: US-09-914-152-1
Perfect score: 1657
Sequence: 1 MAFPKMRLMYICLLVLGALC.....TLLDYWQALENSRGEDCPPV 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1657	100.0	310	21	AAB26749	Human beta-1,3 gal
2	1651	99.6	310	21	AAY94641	Human beta3Gal-T5
3	1016	61.3	192	23	AAE15931	Human 5174397/63-2
4	587	35.4	326	15	AAR57433	Beta-1,3-galactosy
5	569	34.3	422	20	AAY02366	Polypeptide identi
6	555.5	33.5	319	20	AAY06461	Human Dendriac. H
7	555.5	33.5	331	21	AAY66685	Membrane-bound pro
8	555.5	33.5	331	22	AAU29104	Human PRO polypept
9	555.5	33.5	331	22	AAB87543	Human PRO1074. Ho
10	555.5	33.5	331	22	AAB65208	Human PRO1074 (UNQ
11	555.5	33.5	331	23	AAU80224	Human beta1,3-acet
12	550.5	33.2	331	22	AAB93992	Human protein sequ
13	497.5	30.0	193	23	AAE15927	Human 4566767/84-3
14	497.5	30.0	193	23	AAE15928	Mouse 2745735/84-3
15	480	29.0	196	23	AAE15929	Human 4502339_JBC_
16	479	28.9	196	23	AAE15930	Mouse 2745737/157-
17	443.5	26.8	195	23	AAE15933	Mouse 2745739/84-3
18	433.5	26.2	195	23	AAE15932	Human 4502343_JBC_
19	416	25.1	372	19	AAW64558	Human epidermoid c
20	416	25.1	372	22	AAU29167	Human PRO polypept
21	416	25.1	372	22	AAB88404	Human membrane or
22	416	25.1	372	22	AAB49749	Human beta 1,3-N-a
23	416	25.1	372	22	AAB49750	Human beta 1,3-N-a
24	416	25.1	372	23	ABB09716	Amino acid sequenc
25	416	25.1	372	23	AAU11272	Human beta1,3-N-ac
26	402.5	24.3	368	23	AAU11273	Murine beta1,3-N-a
27	396	23.9	378	22	AAE04752	Human beta-1,3-gal
28	396	23.9	378	23	AAE21679	Human galactosyltr
29	396	23.9	378	23	ABB06980	Human beta 1,3-N-a
30	396	23.9	378	23	ABB07522	Human drug metabol
31	394	23.8	401	23	AAE22148	Human TRNFR-10 pro
32	393	23.7	371	20	AAW95171	Protein exhibiting
33	388	23.4	378	20	AAY34171	Human galactosyltr
34	388	23.4	378	22	AAM93908	Human polypeptide,
35	381.5	23.0	574	22	ABB63976	Drosophila melanog
36	375.5	22.7	334	22	ABG20251	Novel human diagno
37	360	21.7	373	22	AAB93536	Human protein sequ
38	360	21.7	397	19	AAW80212	Human cardiac and
39	360	21.7	397	21	AAY84683	Amino acid sequenc
40	360	21.7	397	21	AAY69698	Human Brainiac pro
41	360	21.7	397	22	AAG66118	Human LIG46 polype
42	360	21.7	413	22	AAB49748	Human beta 1,3-N-a
43	359.5	21.7	378	21	AAB07435	The beta-1,3-galac
44	359.5	21.7	402	21	AAB24035	Human PRO4397 prot
45	359.5	21.7	402	23	AAE22142	Human TRNFR-4 prot

ALIGNMENTS

RESULT 1
AAB26749
ID AAB26749 standard; Protein; 310 AA.
XX
AC AAB26749;
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta-1,3 galactose transferase protein sequence.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system.
XX
OS Homo sapiens.
XX
PN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI; 2000-549409/50.
DR N-PSDB; AAA93875.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer -

```
XX
PS   Claim 2; Page 97-98; 123pp; Japanese.
XX
CC   This invention relates to a polypeptide (I) with beta-1,3 galactose
CC   transferase activity, or variants of (I) comprising amino acid additions,
CC   deletions and/or substitutions. Included in the invention is DNA encoding
CC   all or part of (I); expression vectors containing the DNA, host cells
CC   transformed by the vectors; a method for the preparation of the
CC   polypeptide by culture of the transformants or by expression in the milk
CC   of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC   galactose transferase protein transfers galactose by beta-1,3 bonding to
CC   N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC   GlcNAcbeta1-3Galbeta1-4Glc) to give Galbeta1-3GlcNAc. The protein and
CC   DNA encoding it are useful for the treatment and diagnosis of cancer of
CC   the digestive system. The present sequence represents the Beta-1,3
CC   galactose transferase protein.
XX
SQ   Sequence   310 AA;

  Query Match             100.0%;  Score 1657;  DB 21;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-178;
  Matches  310;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60

Qy      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120

Qy     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180

Qy     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240

Qy     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300

Qy     301 NSRGEDCPPV 310
          ||||||||||
Db     301 NSRGEDCPPV 310


RESULT 2
AAY94641
ID   AAY94641 standard; Protein; 310 AA.
XX
AC   AAY94641;
XX
DT   15-AUG-2000  (first entry)
XX
DE   Human beta3Gal-T5 protein sequence.
XX
KW   UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;
KW   Beta3Gal-T5; human; chromosome 21q22.3; galactosylation; glycoprotein;
KW   beta1,3-galactosyl glycosylated saccharide production; glycopeptide.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          13..26
FT                   /note= "Transmembrane domain"
XX
PN   WO200029558-A1.
XX
PD   25-MAY-2000.
XX
PF   11-NOV-1999;   99WO-US26807.
```

— filed before Nov. 29, 2000

```
XX
PR   13-NOV-1998;   98DK-0001483.
XX
PA   (CLAU/) CLAUSEN H.
XX
PI   Clausen H,  Amado M;
XX
DR   WPI; 2000-399728/34.
DR   N-PSDB; AAA27959.
XX
PT   Novel nucleic acid sequence encoding human
PT   UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
PT   useful for obtaining beta 1,3-galactosyl glycosylated saccharides and
PT   glycopeptides or glycoproteins
XX
PS   Claim 23; Fig 1; 74pp; English.
XX
CC   The present invention relates to a nucleic acid sequence encoding
CC   UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
CC   (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at
CC   carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence
CC   represents the human beta3Gal-T5 protein sequence. The beta3Gal-T5 gene
CC   is located on human chromosome 21q22.3. Beta3Gal-T5 is a type II
CC   transmembrane glycoprotein. The invention also relates to the beta3Gal-T5
CC   protein sequence, a nucleic acid vector comprising the beta3Gal-T5
CC   nucleotide sequence, a host cell comprising the vector, and a method for
CC   the production of the beta3Gal-T5 protein from the host cells. The
CC   methods of the invention can be used for recombinant production of
CC   beta3Gal-T5 for use as a catalyst and for recombinant production of
CC   peptides or proteins with appropriate galactosylation. The beta3Gal-T5
CC   protein can be used to obtain beta1,3-galactosyl glycosylated
CC   saccharides, glycopeptides or glycoproteins.
XX
SQ   Sequence   310 AA;

  Query Match             99.6%;  Score 1651;  DB 21;  Length 310;
  Best Local Similarity   99.7%;  Pred. No. 7.1e-178;
  Matches  309;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLV 60

Qy      61 LLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      61 LLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDII 120

Qy     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFF 180

Qy     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240

Qy     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALE 300

Qy     301 NSRGEDCPPV 310
          ||||||||||
Db     301 NSRGEDCPPV 310


RESULT 3
AAE15931
ID   AAE15931 standard; Protein; 192 AA.
XX
AC   AAE15931;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human 5174397/63-284 protein, member of Fringe and Brainiac family.
```

```
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; human.
XX
OS   Homo sapiens.
XX
PN   WO200187321-A2.
XX
PD   22-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-IB01033.
XX
PR   19-MAY-2000; 2000GB-0012216.
XX
PA   (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Cohen S, Brueckner K, Clausen H, Keck B;
XX
DR   WPI; 2002-082940/11.
XX
PT   Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT   treatment of disease caused by epidermal growth factor-like module
PT   containing protein, such as T cell leukemia, breast cancer, stroke and
PT   dementia -
XX
PS   Claim 7; Fig 1; 52pp; English.
XX
CC   The invention relates to the use of a Fringe protein or a Brainiac
CC   protein or a fragment or functional equivalent of a Fringe or a
CC   Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC   proteins have been found to possess glycosyltransferase activity in
CC   transferring sugar residues onto certain proteins of biological
CC   interest, so affecting the binding of effector molecules to these
CC   proteins. Particularly they are useful for transferring a
CC   N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC   whether free or attached to a lipid, carbohydrate or protein. They
CC   act as glycosyltransferase on notch protein (EGF-module containing
CC   protein) family. They are useful in treatment of a disease caused
CC   by epidermal growth factor (EGF)-like module containing protein,
CC   such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC   autosomal dominant arteriopathy with subcortical infarcts,
CC   leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC   proteins are also useful for screening ligands capable of modulating
CC   the activity of the protein. The present sequence is human protein
CC   which is a member of Fringe and Brainiac protein family.
XX
SQ   Sequence   192 AA;

  Query Match             61.3%;  Score 1016;  DB 23;  Length 192;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-106;
  Matches  192;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      63 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 60

Qy     123 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTG 182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTG 120

Qy     183 FLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLE 242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLE 180

Qy     243 DVFVGLCLERLN 254
           ||||||||||||
Db     181 DVFVGLCLERLN 192
```

RESULT 4
AAR57433

```
ID   AAR57433 standard; Protein; 326 AA.
XX
AC   AAR57433;
XX
DT   14-MAR-1995  (first entry)
XX
DE   Beta-1,3-galactosyltransferase.
XX
KW   Beta-1,3-galactosyltransferase; enzyme; saccharide chain; pAMoPRWM1;
KW   KJM-1 cells.
XX
OS   Homo sapiens.
XX
PN   JP06181759-A.
XX
PD   05-JUL-1994.
XX
PF   16-DEC-1992;   92JP-0336436.
XX
PR   16-DEC-1992;   92JP-0336436.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
DR   WPI; 1994-251683/31.
DR   N-PSDB; AAR57433.
XX
PT   Beta-galactosyl-transferase DNA and protein - useful for prodn. of
PT   saccharide chains
XX
PS   Claim 1; Page 22-24; 47pp; Japanese.
XX
CC   This sequence represents a beta-1,3-galactosyltransferase.  This
CC   enzyme can be used to produce physiologically active saccharide
CC   chains and varients, and for improvement of saccharide chains bound
CC   to physiologically active proteins.  The cDNA encoding this protein
CC   represents a fragment of the plasmid pAMoPRWM1 which was cloned in
CC   KJM-1 cells.
XX
SQ   Sequence   326 AA;

  Query Match             35.4%;  Score 587;  DB 15;  Length 326;
  Best Local Similarity   38.6%;  Pred. No. 2e-57;
  Matches  127;  Conservative   52;  Mismatches  108;  Indels   42;  Gaps    7;

Qy      12 CLLVLGALCLYFSMYSLN-----------PFKEQSFVYKK--DGN-----------FL 45
           || ||  :|   ::: |:            ||   :   |   ||            ||
Db       7 CLYVLTVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFL 66

Qy      46 KLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE 105
              | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db      67 INEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVL 126

Qy     106 TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDY 165
            : |:||||   ||| :||:| |:||||||:||: ||  || :| :|||||||:|:|:|
Db     127 NQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDN 186

Qy     166 LTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSG 223
           |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:||
Db     187 LIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSA 245

Qy     224 DVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVCLF 278
           |||  :|  |     : ||||:|||| :| |          |   |     : :|:| :
Db     246 DVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLCRY 297

Qy     279 RRIVACHFIKPRTLLDYWQALENSRGEDC 307
           ||::  | | |  :   |  : : :   |
Db     298 RRVITVHQISPEEMHRIWNDMSSKKHLRC 326
```

RESULT 5
AAY02366
```
ID   AAY02366 standard; Protein; 422 AA.
XX
```

```
AC   AAY02366;
XX
DT   13-JUL-1999  (first entry)
XX
DE   Polypeptide identified by the signal sequence trap method.
XX
KW   Signal sequence trap method; SST method; immunisation; inhibition;
KW   infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW   activin activity; inhibin activity; chemokine activity;
KW   cytokine activity; blood coagulation regulation; agonist; antagonist;
KW   metabolic disorder; hormonal disorder; immune disorder; wound;
KW   severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer.
XX
OS   Homo sapiens.
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
XX
DR   WPI; 1999-277254/23.
DR   N-PSDB; AAX35710, AAX35711.
XX
PT   Polypeptides identified by the signal sequence trap method from a
PT   human cDNA library
XX
PS   Claim 1; Page 106-108; 281pp; Japanese.
XX
CC   AAY02358-84 represent novel polypeptides which are identified from a
CC   human placental cDNA library by the signal sequence trap (SST) method.
CC   The polypeptides are encoded by the cDNA seqeuences in AAX35694-X35747.
CC   The polypeptides have a broad range of physiological activity, including
CC   immunisation against and inhibition of infections, allergies and cancer;
CC   regulation of tissue formation and repair; activin/inhibin activity;
CC   chemokine/cytokine activity; blood coagulation regulation; and
CC   receptor/ligand agonist or antagonist activity. The polypeptides can
CC   be used for prevention and treatment of disorders including infections
CC   by bacteria, yeasts and viruses (including HIV) and protozoa; metabolic
CC   and hormonal disorders; immune disorders (including severe combined
CC   immunodeficiency (SCID) and AIDS; thrombosis; cancer; and traumatic or
CC   surgical wounds.
XX
SQ   Sequence   422 AA;

  Query Match             34.3%;  Score 569;  DB 20;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 3.1e-55;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
            :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405


RESULT 6
AAY06461
ID   AAY06461 standard; Protein; 319 AA.
XX
AC   AAY06461;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Human Dendriac.
XX
KW   Dendriac; Brainiac-2; human; Notch; immune disorder;
KW   neurological disorder; diagnosis; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /note= "signal peptide"
FT   Protein         26..319
FT                   /note= "mature protein"
FT   Modified-site   60
FT                   /note= "N-glycosylated"
FT   Modified-site   142
FT                   /note= "N-glycosylated"
FT   Modified-site   186
FT                   /note= "N-glycosylated"
FT   Modified-site   200
FT                   /note= "N-glycosylated"
FT   Modified-site   314
FT                   /note= "N-glycosylated"
FT   Domain          82..90
FT                   /label= CD-I
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          100..109
FT                   /label= CD-II
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          120..153
FT                   /label= CD-III
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          168..182
FT                   /label= CD-IV
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          199..204
FT                   /label= CD-V
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          217..225
FT                   /label= CD-VI
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          234..242
FT                   /label= CD-VII
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          250..261
FT                   /label= CD-VIII
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          274..282
FT                   /label= CD-IX
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
FT   Domain          285..291
FT                   /label= CD-X
FT                   /note= "conserved domain, shows high identity
FT                        to Brainiac-3 and Drosophila Brainiac"
```

```
FT   Domain          297..308
FT                   /label= CD-XI
FT                   /note= "conserved domain, shows high identity
FT                         to Brainiac-3 and Drosophila Brainiac"
FT   Peptide         55..63
FT                   /note= "epitope-bearing peptide"
FT   Peptide         75..84
FT                   /note= "epitope-bearing peptide"
FT   Peptide         89..95
FT                   /note= "epitope-bearing peptide"
FT   Peptide         233..238
FT                   /note= "epitope-bearing peptide"
FT   Peptide         249..255
FT                   /note= "epitope-bearing peptide"
FT   Peptide         271..277
FT                   /note= "epitope-bearing peptide"
FT   Peptide         298..302
FT                   /note= "epitope-bearing peptide"
XX
PN   WO9931116-A1.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-US27049.
XX
PR   17-NOV-1998;   98US-0108928.
PR   18-DEC-1997;   97US-0068006.
PR   12-MAR-1998;   98US-0077687.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ebner R,  Endress GA,  Florence KA,  Rosen CA,  Ruben SM;
PI   Soppet DR,  Yu G;
XX
DR   WPI; 1999-430031/36.
DR   N-PSDB; AAX87192, AAX87193.
XX
PT   Human proteins for treating and detecting immune and neurological
PT   disorders
XX
PS   Claim 22; Page 112-113; 132pp; English.
XX
CC   The present sequence represents human Dendriac, also called
CC   Brainiac-2, a novel member of the Brainiac family.  The sequence
CC   was deduced from Dendriac cDNA (see AAX87192 and AAX87193) initially
CC   discovered in a dendritic cell library.  Additional clones of the
CC   same gene were identified in NTERA2 cell, adult pulmonary tissue,
CC   salivary gland, ovary, Caco-2 colon adenocarcinoma, smooth muscle,
CC   cerebellum, 8-week-old human mrbyo, haemagiopericytoma, amygdala,
CC   substantia nigra and whole brain cDNA libraries.  Homology is
CC   shown to Brainiac-3 (see AAY06462) and to Drosophila Brainiac.
CC   Recombinant Dendriac polypeptides, especially the mature protein
CC   or epitope-bearing portions of it, can be expressed in host cells
CC   using claimed recombinant vectors.  Dendriac and Brainiac proteins
CC   are members of the Notch family of proteins involved in the control
CC   of cell division.  The proteins may be used to detect and/or treat
CC   immune system and neurological disorders, and to raise specific
CC   antibodies.
XX
SQ   Sequence   319 AA;

  Query Match             33.5%;  Score 555.5;  DB 20;  Length 319;
  Best Local Similarity   36.2%;  Pred. No. 7e-54;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |            | |   :|::| :|     ::|
Db       3 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 62

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :       ::
Db      63 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 122

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
             |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     123 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 182

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     183 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 242

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
             :   |  || |||:||:||  | : :        ||   :   ||  ||::| |   :
Db     243 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 302

Qy     291 TLLDYWQAL 299
             :: :|| :
Db     303 EIITFWQVM 311


RESULT 7
AAY66685
ID   AAY66685 standard; protein; 331 AA.
XX
AC   AAY66685;
XX
DT   05-APR-2000  (first entry)
XX
DE   Membrane-bound protein PRO1074.
XX
KW   Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW   pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS   Homo sapiens.
XX
PN   WO9963088-A2.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999;   99WO-US12252.
XX
PR   02-JUN-1998;   98US-0087607.
PR   02-JUN-1998;   98US-0087609.
PR   02-JUN-1998;   98US-0087759.
PR   03-JUN-1998;   98US-0087827.
PR   04-JUN-1998;   98US-0088021.
PR   04-JUN-1998;   98US-0088025.
PR   04-JUN-1998;   98US-0088028.
PR   04-JUN-1998;   98US-0088029.
PR   04-JUN-1998;   98US-0088030.
PR   04-JUN-1998;   98US-0088033.
PR   04-JUN-1998;   98US-0088326.
PR   05-JUN-1998;   98US-0088167.
PR   05-JUN-1998;   98US-0088202.
PR   05-JUN-1998;   98US-0088212.
PR   05-JUN-1998;   98US-0088217.
PR   09-JUN-1998;   98US-0088655.
PR   10-JUN-1998;   98US-0088722.
PR   10-JUN-1998;   98US-0088730.
PR   10-JUN-1998;   98US-0088734.
PR   10-JUN-1998;   98US-0088738.
PR   10-JUN-1998;   98US-0088740.
PR   10-JUN-1998;   98US-0088741.
PR   10-JUN-1998;   98US-0088742.
PR   10-JUN-1998;   98US-0088810.
PR   10-JUN-1998;   98US-0088811.
PR   10-JUN-1998;   98US-0088824.
PR   10-JUN-1998;   98US-0088825.
PR   10-JUN-1998;   98US-0088826.
PR   11-JUN-1998;   98US-0088858.
PR   11-JUN-1998;   98US-0088861.
PR   11-JUN-1998;   98US-0088863.
PR   11-JUN-1998;   98US-0088876.
PR   12-JUN-1998;   98US-0089090.
PR   12-JUN-1998;   98US-0089105.
PR   16-JUN-1998;   98US-0089440.
```

PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 16-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
DR WPI; 2000-072883/06.
DR N-PSDB; AAZ65022.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
PS claim 12; Fig 137; 822pp; English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
SQ Sequence 331 AA;

Query Match 33.5%; Score 555.5; DB 21; Length 331;
Best Local Similarity 36.2%; Pred. No. 7.4e-54;
Matches 112; Conservative 57; Mismatches 125; Indels 15; Gaps 2;

```
Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |            | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194
```

```
Qy    171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
          |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db    195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy    231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
          :   |  || |||:||:||  | : :        ||   :   ||  ||::| |    :
Db    255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy    291 TLLDYWQAL 299
            :: :|| :
Db    315 EIITFWQVM 323


RESULT 8
AAU29104
ID   AAU29104 standard; Protein; 331 AA.
XX
AC   AAU29104;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human PRO polypeptide sequence #81.
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW   dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200168848-A2.
XX
PD   20-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06520.
XX
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   03-MAR-2000; 2000US-187202P.
PR   06-MAR-2000; 2000US-186968P.
PR   14-MAR-2000; 2000US-189320P.
PR   14-MAR-2000; 2000US-189328P.
PR   15-MAR-2000; 2000WO-US06884.
PR   21-MAR-2000; 2000US-190828P.
PR   21-MAR-2000; 2000US-191007P.
PR   21-MAR-2000; 2000US-191048P.
PR   21-MAR-2000; 2000US-191314P.
PR   28-MAR-2000; 2000US-192655P.
PR   29-MAR-2000; 2000US-193032P.
PR   29-MAR-2000; 2000US-193053P.
PR   30-MAR-2000; 2000WO-US08439.
PR   04-APR-2000; 2000US-194449P.
PR   04-APR-2000; 2000US-194647P.
PR   11-APR-2000; 2000US-195975P.
PR   11-APR-2000; 2000US-196000P.
PR   11-APR-2000; 2000US-196187P.
PR   11-APR-2000; 2000US-196690P.
PR   11-APR-2000; 2000US-196820P.
PR   18-APR-2000; 2000US-198121P.
PR   18-APR-2000; 2000US-198585P.
PR   25-APR-2000; 2000US-199397P.
PR   25-APR-2000; 2000US-199550P.
PR   25-APR-2000; 2000US-199654P.
PR   03-MAY-2000; 2000US-201516P.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   05-JUN-2000; 2000US-209832P.
PR   28-JUL-2000; 2000WO-US20710.
PR   22-AUG-2000; 2000US-0644848.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI   Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2001-602746/68.
DR   N-PSDB; AAS46005.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
XX
PS   Claim 11; Fig 162; 774pp; English.
XX
CC   Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC   The PRO polypeptides and their associated nucleic acids can be used to
CC   detect the presence of a tumour in a mammal by comparing the level of
CC   expression of a PRO polypeptide in a test sample of cells from the animal
CC   and a control sample of normal cells, whereby a higher level of
CC   expression in the test sample indicates the presence of a tumour in the
CC   mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC   and rabbits but are preferably human. The polypeptides can be used to
CC   stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC   when contacted with it. A specific polypeptide can be used to stimulate
CC   the proliferation or differentiation of chondrocyte cells. The PRO
CC   proteins can be used to determine the presence of tumours and also
CC   susceptibility to tumour development, particularly adrenal, lung, colon,
CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC   can be used for genetic analysis of individuals with genetic disorders.
XX
SQ   Sequence   331 AA;

  Query Match           33.5%;  Score 555.5;  DB 22;  Length 331;
  Best Local Similarity  36.2%;  Pred. No. 7.4e-54;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy      6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
          :| :   ||:|  |  :   |   |            | |   :|::| :|    ::|
Db     15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy     54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
            ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db     75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy    111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
          |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db    135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy    171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
          |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db    195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy    231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
          :   |  || |||:||:||  | : :        ||   :   ||  ||::| |    :
Db    255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy    291 TLLDYWQAL 299
            :: :|| :
Db    315 EIITFWQVM 323


RESULT 9
AAB87543
ID   AAB87543 standard; Protein; 331 AA.
XX
AC   AAB87543;
XX
DT   15-MAY-2001  (first entry)
```

```
XX
DE   Human PRO1074.
XX
KW   Human; PRO protein; mapping.
XX
OS   Homo sapiens.
XX
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000; 2000US-0175481.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   01-MAR-2000; 2000WO-US05601.
PR   03-MAR-2000; 2000US-0187202.
PR   25-APR-2000; 2000US-0199397.
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   N-PSDB; AAF92075.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 12; Fig 36; 278pp; English.
XX
CC   The present sequence is a human PRO polypeptide (secreted and
CC   transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC   anti-PRO antibodies are useful for preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the PRO protein,
CC   agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC   employed as molecular weight markers for protein electrophoresis. The PRO
CC   coding sequence has applications in molecular biology, including use as
CC   hybridisation probes, and in chromosome and gene mapping.
XX
SQ   Sequence   331 AA;

  Query Match             33.5%;  Score 555.5;  DB 22;  Length 331;
  Best Local Similarity   36.2%;  Pred. No. 7.4e-54;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |             | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :       ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
           :   |  || |||:||:||  | : :        ||   :   ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
             :: :|| :
Db     315 EIITFWQVM 323


RESULT 10
AAB65208
ID   AAB65208 standard; Protein; 331 AA.
XX
AC   AAB65208;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO1074 (UNQ531) protein sequence SEQ ID NO:209.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   N-PSDB; AAF44168.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 12; Fig 137; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
```

```
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence   331 AA;

  Query Match             33.5%;  Score 555.5;  DB 22;  Length 331;
  Best Local Similarity   36.2%;  Pred. No. 7.4e-54;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |  |              | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |  :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
             :   |  || |||:||:||  | : :        ||   :  ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
           :: :|| :
Db     315 EIITFWQVM 323


RESULT 11
AAU80224
ID   AAU80224 standard; Protein; 331 AA.
XX
AC   AAU80224;
XX
DT   30-JUL-2002  (first entry)
XX
DE   Human beta1,3-acetylgalactosamine transferase.
XX
KW   Human; enzyme; beta1,3-acetylgalactosamine transferase;
KW   Gb4 sugar chain.
XX
OS   Homo sapiens.
XX
PN   JP2002085069-A.
XX
PD   26-MAR-2002.
XX
PF   08-SEP-2000; 2000JP-0273835.
XX
PR   08-SEP-2000; 2000JP-0273835.
XX
PA   (SEGK ) SEIKAGAKU KOGYO CO LTD.
PA   (FURU/) FURUKAWA K.
XX
DR   WPI; 2002-378274/41.
DR   N-PSDB; ABK51201.
XX
PT   Preparation of beta-1,3-acetylgalactosamine transferase -
XX
PS   Claim 3; Page 13-14; 15pp; Japanese.
XX
CC   The invention relates to the preparation of beta-1,3-acetylgalactosamine
CC   (a Gb4 sugar chain molecule) comprising introducing a DNA appearing as
CC   ABK51201 (S1) or a DNA hybridising with S1 or a base sequence
CC   complementary to the base sequence or part of these base sequences under
CC   a stringent condition into a cell and growing the cell to express beta-1,
CC   3-acetylgalactosamine transferase and collecting it. Also included is the
CC   preparation of Gb4 sugar chain comprising contacting a polypeptide
CC   comprising the beta-1,3-acetylgalactosamine transferase protein appearing
CC   as AAU80224 (A) or a polypeptide consisting of an amino acid sequence in
CC   which at least one amino acid is replaced, deleted, inserted or
CC   transferred in the amino acid sequence (A) and having enzymatic activity
CC   transferring N-acetylgalactosamine residue from an N-acetylgalactosamine
CC   donor to the C3 site of the galactose residue in Gb3 sugar chain which is
CC   the receptor with an N-acetylgalactosamine donor and Gb3 sugar chain.
CC   The present sequence represents beta-1,3-acetylgalactosamine transferase.
XX
SQ   Sequence   331 AA;

  Query Match             33.5%;  Score 555.5;  DB 23;  Length 331;
  Best Local Similarity   36.2%;  Pred. No. 7.4e-54;
  Matches  112;  Conservative   57;  Mismatches  125;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |  |              | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
             ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |  :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:  :||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
             :   |  || |||:||:||  | : :        ||   :  ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
           :: :|| :
Db     315 EIITFWQVM 323


RESULT 12
AAB93992
ID   AAB93992 standard; Protein; 331 AA.
XX
AC   AAB93992;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14093.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
```

```
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14093; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   331 AA;

  Query Match             33.2%;  Score 550.5;  DB 22;  Length 331;
  Best Local Similarity   36.2%;  Pred. No. 2.7e-53;
  Matches  112;  Conservative   56;  Mismatches  126;  Indels   15;  Gaps    2;

Qy       6 MRLMYICLLVLGALCLYFSMYSLNP-----------FKEQSFVYKKDGNFLKLPDTDCR 53
           :| :   ||:|  |  :   |   |            | |   :|::| :|     ::|
Db      15 LRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCS 74

Qy      54 QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVD 110
              ||||:||||    :  | ||| |||:::   | :: ||||||  :   :      ::
Db      75 HQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLE 134

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL 170
           |   :||||::|||| | ||||||:|   ||  ||| | :|||||:|:|||   | : |
Db     135 DEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYL 194

Qy     171 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY 230
           |  | : :||||:  :: :  |  : |  :|  |||:   ||:||| ||: | |:  ::|
Db     195 LNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVSPPYCSGLGYIMSRDLVPRIY 254

Qy     231 NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPR 290
             :   |  || |||:||:||  | : :        ||   :   ||  ||::| |    :
Db     255 EMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSK 314

Qy     291 TLLDYWQAL 299
             :: :|| :
Db     315 EIITFWQVM 323
```

```
RESULT 13
AAE15927
ID   AAE15927 standard; Protein; 193 AA.
XX
AC   AAE15927;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human 4566767/84-303 protein, member of Fringe and Brainiac family.
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; human.
XX
OS   Homo sapiens.
XX
PN   WO200187321-A2.
XX
PD   22-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-IB01033.
XX
PR   19-MAY-2000; 2000GB-0012216.
XX
PA   (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Cohen S,  Brueckner K,  Clausen H,  Keck B;
XX
DR   WPI; 2002-082940/11.
XX
PT   Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT   treatment of disease caused by epidermal growth factor-like module
PT   containing protein, such as T cell leukemia, breast cancer, stroke and
PT   dementia -
XX
PS   Claim 7; Fig 1; 52pp; English.
XX
CC   The invention relates to the use of a Fringe protein or a Brainiac
CC   protein or a fragment or functional equivalent of a Fringe or a
CC   Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC   proteins have been found to possess glycosyltransferase activity in
CC   transferring sugar residues onto certain proteins of biological
CC   interest, so affecting the binding of effector molecules to these
CC   proteins. Particularly they are useful for transferring a
CC   N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC   whether free or attached to a lipid, carbohydrate or protein. They
CC   act as glycosyltransferase on notch protein (EGF-module containing
CC   protein) family. They are useful in treatment of a disease caused
CC   by epidermal growth factor (EGF)-like module containing protein,
CC   such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC   autosomal dominant arteriopathy with subcortical infarcts,
CC   leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC   proteins are also useful for screening ligands capable of modulating
CC   the activity of the protein. The present sequence is human protein
CC   which is a member of Fringe and Brainiac protein family.
XX
SQ   Sequence   193 AA;

  Query Match             30.0%;  Score 497.5;  DB 23;  Length 193;
  Best Local Similarity   49.7%;  Pred. No. 1.2e-47;
  Matches   96;  Conservative   34;  Mismatches   60;  Indels    3;  Gaps    2;

Qy      63 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 122
           ::::||:   | |||:||| |   || :: | ||||  :     : |:||||   ||| :
Db       1 ISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVLNQMVEQESQIFHDIIVE 60

Qy     123 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNR--TTRFF 180
           ||:| |:|||||:||: ||  || :| :|||||||:|:|:| |   ||| :     |:|
Db      61 DFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDNLIYRLLKPSTKPRRRYF 120

Qy     181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
           ||:: :|  |||   |||:: :  ||   ||||||||||:|| |||  :|  |     :
```

```
Db    121 TGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLH 179

Qy    241 LEDVFVGLCLERL 253
          ||||:||||| :|
Db    180 LEDVYVGLCLRKL 192


RESULT 14
AAE15928
ID   AAE15928 standard; Protein; 193 AA.
XX
AC   AAE15928;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Mouse 2745735/84-303 protein, member of Fringe and Brainiac family.
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; mouse.
XX
OS   Mus sp.
XX
PN   WO200187321-A2.
XX
PD   22-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-IB01033.
XX
PR   19-MAY-2000; 2000GB-0012216.
XX
PA   (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Cohen S,  Brueckner K,  Clausen H,  Keck B;
XX
DR   WPI; 2002-082940/11.
XX
PT   Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT   treatment of disease caused by epidermal growth factor-like module
PT   containing protein, such as T cell leukemia, breast cancer, stroke and
PT   dementia -
XX
PS   Claim 7; Fig 1; 52pp; English.
XX
CC   The invention relates to the use of a Fringe protein or a Brainiac
CC   protein or a fragment or functional equivalent of a Fringe or a
CC   Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC   proteins have been found to possess glycosyltransferase activity in
CC   transferring sugar residues onto certain proteins of biological
CC   interest, so affecting the binding of effector molecules to these
CC   proteins. Particularly they are useful for transferring a
CC   N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC   whether free or attached to a lipid, carbohydrate or protein. They
CC   act as glycosyltransferase on notch protein (EGF-module containing
CC   protein) family. They are useful in treatment of a disease caused
CC   by epidermal growth factor (EGF)-like module containing protein,
CC   such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC   autosomal dominant arteriopathy with subcortical infarcts,
CC   leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC   proteins are also useful for screening ligands capable of modulating
CC   the activity of the protein. The present sequence is mouse protein
CC   which is a member of Fringe and Brainiac protein family.
XX
SQ   Sequence   193 AA;

  Query Match             30.0%;  Score 497.5;  DB 23;  Length 193;
  Best Local Similarity   49.7%;  Pred. No. 1.2e-47;
  Matches   96;  Conservative   34;  Mismatches   60;  Indels    3;  Gaps    2;

Qy     63 VTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQK 122
          ::::||:   | |||:||| |   || :: | ||||  :     : |:||||   ||| :
Db      1 ISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVLNQMVEQESQIFHDIIVE 60

Qy    123 DFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNR--TTRFF 180
          ||:| |:|||||:||: ||  || :| :|||||:|:|:| |   ||| :     |:|
Db     61 DFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDNLIYRLLKPSTKPRRRYF 120

Qy    181 TGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIK 240
          ||:: :|  |||   |||:: :  ||   ||||||||||:|| |||  :|  |     :
Db    121 TGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLH 179

Qy    241 LEDVFVGLCLERL 253
          ||||:||||| :|
Db    180 LEDVYVGLCLRKL 192


RESULT 15
AAE15929
ID   AAE15929 standard; Protein; 196 AA.
XX
AC   AAE15929;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human 4502339_JBC_1/157-364 protein member of Fringe and Brainiac family.
XX
KW   Fringe protein; Brainiac protein; glycosyltransferase; EGF; mannose;
KW   N-acetylglucosamine; fucose; epidermal growth factor; T cell leukaemia;
KW   breast cancer; stroke; dementia; leucoencephalopathy; Alagille syndrome;
KW   notch protein; EGF-module containing protein; human.
XX
OS   Homo sapiens.
XX
PN   WO200187321-A2.
XX
PD   22-NOV-2001.
XX
PF   21-MAY-2001; 2001WO-IB01033.
XX
PR   19-MAY-2000; 2000GB-0012216.
XX
PA   (EUMO-) EURO MOLECULAR BIOLOGY LAB.
XX
PI   Cohen S,  Brueckner K,  Clausen H,  Keck B;
XX
DR   WPI; 2002-082940/11.
XX
PT   Use of Fringe protein or Brainiac protein as glycosyltransferase and
PT   treatment of disease caused by epidermal growth factor-like module
PT   containing protein, such as T cell leukemia, breast cancer, stroke and
PT   dementia -
XX
PS   Claim 7; Fig 1; 52pp; English.
XX
CC   The invention relates to the use of a Fringe protein or a Brainiac
CC   protein or a fragment or functional equivalent of a Fringe or a
CC   Brainiac protein, as glycosyltransferase. Fringe and Brainiac
CC   proteins have been found to possess glycosyltransferase activity in
CC   transferring sugar residues onto certain proteins of biological
CC   interest, so affecting the binding of effector molecules to these
CC   proteins. Particularly they are useful for transferring a
CC   N-acetylglucosamine moiety onto a fucose or a mannose substrate,
CC   whether free or attached to a lipid, carbohydrate or protein. They
CC   act as glycosyltransferase on notch protein (EGF-module containing
CC   protein) family. They are useful in treatment of a disease caused
CC   by epidermal growth factor (EGF)-like module containing protein,
CC   such as T cell leukaemia, breast cancer, stroke, dementia, cerebral
CC   autosomal dominant arteriopathy with subcortical infarcts,
CC   leucoencephalopathy and Alagille syndrome. Fringe and Brainiac
CC   proteins are also useful for screening ligands capable of modulating
CC   the activity of the protein. The present sequence is human protein
CC   which is a member of Fringe and Brainiac protein family.
XX
SQ   Sequence   196 AA;
```

Query Match 29.0%; Score 480; DB 23; Length 196;
Best Local Similarity 48.1%; Pred. No. 1.2e-45;
Matches 91; Conservative 36; Mismatches 58; Indels 4; Gaps 2;

```
Qy      69 QLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEVDQESQRHGDIIQKDFLD 126
           |:  | |||||| | :  | |:   |||| :   :    : : :||::: ||||:::||
Db       7 QIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAILEESRQYHDIIQQEYLD 66

Qy     127 VYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTR--FFTGFL 184
            |||||:||:||: ||  :||   :|||||||||:| :||   ||| :   |  :|||:|
Db      67 TYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINRLLKPDLPPRHNYFTGYL 126

Qy     185 KLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDV 244
             | |   |||::    || :||| ||||||||||||:| ::: ||  :  : ||||
Db     127 MRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAEKIFKVSLGIRRLHLEDV 186

Qy     245 FVGLCLERL 253
           :||:|| :|
Db     187 YVGICLAKL 195
```

Search completed: April 14, 2003, 12:10:27
Job time : 77 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 12:08:21 ; Search time 46 Seconds
(without alignments)
647.862 Million cell updates/sec

Title: US-09-914-152-1
Perfect score: 1657
Sequence: 1 MAFPKMRLMYICLLVLGALC.....TLLDYWQALENSRGEDCPPV 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	295.5	17.8	684	2	F86394	protein T24P13.20
2	267	16.1	657	2	G86397	protein T7N9.18 [i
3	265	16.0	739	2	T10648	hypothetical prote
4	243.5	14.7	642	2	D96777	hypothetical prote
5	233.5	14.1	262	2	T20203	hypothetical prote
6	223.5	13.5	384	2	T20446	hypothetical prote
7	219.5	13.2	507	2	T18639	hypothetical prote
8	209.5	12.6	345	2	T20031	hypothetical prote
9	200	12.1	357	2	T20029	hypothetical prote
10	199	12.0	376	2	T24925	hypothetical prote
11	192	11.6	332	2	T20028	hypothetical prote
12	189.5	11.4	376	2	T20876	hypothetical prote
13	171.5	10.4	398	2	B86353	protein F2E2.6 [im
14	160	9.7	353	2	C96573	protein F12M16.19
15	158.5	9.6	253	2	T22387	hypothetical prote
16	155.5	9.4	325	2	T24762	hypothetical prote
17	146.5	8.8	409	2	A84733	hypothetical prote
18	143.5	8.7	404	2	A86186	hypothetical prote
19	141.5	8.5	359	2	F87937	protein F14B6.6 [i
20	141.5	8.5	385	2	T20879	hypothetical prote
21	140.5	8.5	338	2	T24743	hypothetical prote
22	139.5	8.4	399	2	H86453	avr9 homolog F9L11
23	139	8.4	406	2	T04817	hypothetical prote
24	134	8.1	395	2	A86458	probasble elicitor
25	132.5	8.0	348	2	T31918	hypothetical prote
26	126.5	7.6	334	2	T32256	hypothetical prote
27	122.5	7.4	401	2	A86251	hypothetical prote
28	109.5	6.6	333	2	T02614	hypothetical prote
29	105.5	6.4	378	2	T39790	hypothetical prote
30	105.5	6.4	753	2	T24745	hypothetical prote
31	105.5	6.4	1049	2	S67613	vacuolar carboxype
32	104	6.3	414	2	B96808	protein F28K19.2 [
33	101	6.1	349	2	T24744	hypothetical prote
34	100	6.0	845	2	T46164	nodulin / glutamat
35	98	5.9	660	2	T22794	hypothetical prote
36	95.5	5.8	347	2	D71690	hypothetical prote
37	94.5	5.7	3724	2	T18427	hypothetical prote
38	93.5	5.6	336	2	S70708	site-specific DNA-
39	93	5.6	368	2	T15096	hypothetical prote
40	92.5	5.6	279	2	T20443	hypothetical prote
41	90.5	5.5	338	2	T49998	hypothetical prote
42	90.5	5.5	1314	2	S19488	probable membrane
43	90	5.4	482	2	F82828	phosphomannose iso
44	90	5.4	786	2	T08664	Toll protein-like
45	88.5	5.3	579	2	S46757	hypothetical prote

ALIGNMENTS

RESULT 1
F86394
protein T24P13.20 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: F86394
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alo
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86394
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-684 <STO>
A;Cross-references: GB:AE005172; NID:g9295733; PIDN:AAF87039.1; GSPDB:GN00141
C;Genetics:
A;Gene: T24P13.20
A;Map position: 1

Query Match 17.8%; Score 295.5; DB 2; Length 684;
Best Local Similarity 26.3%; Pred. No. 1.4e-18;
Matches 92; Conservative 58; Mismatches 145; Indels 55; Gaps 10;

```
Qy       9 MYICLLVLGALCLYFSM-------YSLNPFKEQSFVYKKDGNFLK-----LP-------- 48
           :|| ||:|      ||:        :| |:         |   :      ||
Db     335 LYIFLLLLAIFHQMFSVSYGFFVFQTLEPWLVSEIRITGDFRLISILASGLPTSEESEHV 394

Qy      49 -DTDCRQTPPF-------LVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT 100
            | :  ::|          ||: | |:       |||:|:|| :   |:  ::   | :|
Db     395 VDLEALKSPTLSPLRPLDLVIGVFSTANNFKRRMAVRRTWMQYDDVRSGRVAVRFFVGLH 454

Qy     101 SSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTM----MGIEWVHRFC---------- 146
           |     |:  |:: :||:    |:| |  :: ||:     |:  :: |
Db     455 KSPLVNLELWNEARTYGDVQLMPFVDYYSLISWKTLAICIFGLSSLYCFASLTSYFKQTE 514

Qy     147 -PQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVS---- 201
             | |:|||| | |: || :   |   | |     | :  :  ||| | |||:|
Db     515 VDSAKFIMKTDDDAFVRVDEVLLSLSMTNNTRGLIYGLINSDSQPIRNPDSKWYISYEFL 574

Qy     202 --KSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSK--SVPYIKLEDVFVGLCLERLNIRL 257
             | |:| ::|||:  | ||: | |:|  |  : |  ::   ||||| :|: :  |
Db     575 LKKQEWPEEKYPPWAHGPGYIVSRDIAESVGKLFKEGNLKMFKLEDVAMGIWIAELTKHG 634

Qy     258 EELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDC 307
           | | :    | :    |   :|| |:  |  :   |:  : ::   |    .
```

```
Db    635 LEPHYEN---DGRIISDGCKDGYVVA-HYQSPAEMTCLWRKYQETKRSLC 680


RESULT 2
G86397
protein T7N9.18 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86397
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86397
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-657 <STO>
A;Cross-references: GB:AE005172; NID:g8778858; PIDN:AAF79857.1; GSPDB:GN00141
C;Genetics:
A;Gene: T7N9.18
A;Map position: 1

  Query Match             16.1%;  Score 267;  DB 2;  Length 657;
  Best Local Similarity   27.5%;  Pred. No. 5e-16;
  Matches   67;  Conservative   51;  Mismatches  102;  Indels   24;  Gaps    7;

Qy      73 RMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLT 132
           |||:|::| ::::|:  ::   | :   :      :: :|::  |||:   ::| |  :
Db     425 RMAVRKSWMQQKLVRSSKVVARFFVALHARKEVNVDLKKEAEYFGDIVIVPYMDHYDLVV 484

Qy     133 LKTM----MGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNE 188
           |||:     |:  |      | :||| | | |: || : :   |       : | :  |
Db     485 LKTVAICEYGVNTV-----AAKYVMKCDDDTFVRVDAVIQEAEKVKGRESLYIGNINFNH 539

Qy     189 FPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN--VSKSVPYIKLEDVFV 246
           |:|    || |:  |:| : |||: :| ||: | |||  : :   | :   |:||| :
Db     540 KPLRT--GKWAVTFEEWPEEYYPPYANGPGYILSYDVAKFIVDDFEQKRLRLFKMEDVSM 597

Qy     247 GLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRI---VACHFIKPRTLLDYWQALENSR 303
           |: :|: |      ::|      |:|  | |  |      |:  || ::  |  |:
Db     598 GMWVEKFN------ETRPVAVVHSLKF--CQFGCIEDYFTAHYQSPRQMICMWDKLQRLG 649

Qy     304 GEDC 307
           |
Db     650 KPQC 653


RESULT 3
T10648
hypothetical protein T13K14.220 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C;Accession: T10648
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10648
A;Molecule type: DNA
A;Residues: 1-739 <BEV>
A;Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.220
A;Experimental source: cultivar Columbia; BAC clone T13K14
C;Genetics:
A;Gene: ATSP:T13K14.220
A;Map position: 4
A;Introns: 59/3; 86/1; 358/1; 438/3; 535/3; 578/3; 634/3; 670/3

  Query Match             16.0%;  Score 265;  DB 2;  Length 739;
  Best Local Similarity   26.8%;  Pred. No. 8.8e-16;
  Matches   79;  Conservative   54;  Mismatches  138;  Indels   24;  Gaps    9;

Qy      18 ALCLYFSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPF-LVLLVTSSHKQLAERMAI 76
           |  |  |  | :|  :::  :  :     || |      || | : | |:    :||||:
Db     461 ATSLSTSHPSFSP--QKAIEFSSEWKAPPLPGT------PFRLFMGVLSATNHFSERMAV 512

Qy      77 RQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTM 136
           |:|| :   :|   :   | :           : :|::  |||:   |:| |  : |||:
Db     513 RKTWMQHPSIKSSDVVARFFVALNPRKEVNAMLKKEAEYFGDIVILPFMDRYELVVLKTI 572

Qy     137 MGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFS 196
           |: :   |   ::|| | | || |: : : :   :      : | | |   |:|
Db     573 AICEFGNVTAP---YIMKCDDDTFIRVESILKQIDGVSPEKSLYMGNLNLRHRPLRT--G 627

Qy     197 KWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQV--YNVSKSVPYIKLEDVFVGLCLERLN 254
           || |:  |:|   |||: :| ||: | ::|  :   |    :   |:||| :|| :|: |
Db     628 KWTVTWEEWPEAVYPPYANGPGYIISSNIAKYIVSQNSRHKLRLFKMEDVSMGLWVEQFN 687

Qy     255 IRLEELHSQPTFFPGGLRFSV--CLFRRIVACHFIKPRTLLDYWQALENSRGEDC 307
             :    || :    :|   |      | |:  |  ::  |  |   | : |
Db     688 ASM-----QPVEYSHSWKFCQYGCTLNYYTA-HYQSPSQMMCLWDNLLKGRPQCC 736


RESULT 4
D96777
hypothetical protein F25A4.23 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96777
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-642 <STO>
A;Cross-references: GB:AE005173; NID:g5882743; PIDN:AAD55296.1; GSPDB:GN00141
C;Genetics:
A;Gene: F25A4.23
A;Map position: 1

  Query Match             14.7%;  Score 243.5;  DB 2;  Length 642;
  Best Local Similarity   24.9%;  Pred. No. 6.5e-14;
  Matches   65;  Conservative   51;  Mismatches  116;  Indels   29;  Gaps    8;

Qy      58 FLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHG 117
           |: :|   :|   :||||:|::| :  ::   ::   | :          |: :|::  |
Db     396 FIGILSAGNH--FSERMAVRKSWMQHVLITSAKVVARFFVALHGRKEVNVELKKEAEYFG 453

Qy     118 DIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAA------FVMKTDSDMFINVDYLTELLL 171
           ||:   ::| |  : |||:         |   |      ::|| | | |: :  :   :
Db     454 DIVLVPYMDSYDLVVLKTV-------AICEHGALAFSAKYIMKCDDDTFVKLGAVINEVK 506

Qy     172 KKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN 231
           |       : | :    |:|    || |:  |:| : |||: :| ||| | |:|  : :
Db     507 KVPEGRSLYIGNMNYYHKPLRG--GKWAVTYEEWPEEDYPPYANGPGYVLSSDIARFIVD 564

Qy     232 VSK--SVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLF---RRIVACHF 286
           :    :  |:||| ||: :|       :  : |  :   |||  | |          |:
Db     565 KFERHKLRLFKMEDVSVGMWVEHF-----KNTTNPVDYRHSLRF--CQFGCVENYYTAHY 617

Qy     287 IKPRTLLDYWQALENSRGEDC 307
           || ::  |  |      :|
```

```
Db    618 QSPRQMICLWDKLLRQNKPEC 638


RESULT 5
T20203
hypothetical protein C54C8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20203
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19237
A;Accession: T20203
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-262 <WIL>
A;Cross-references: EMBL:Z83102; PIDN:CAB05465.1; GSPDB:GN00019; CESP:C54C8.3
A;Experimental source: clone C54C8
C;Genetics:
A;Gene: CESP:C54C8.3
A;Map position: 1
A;Introns: 19/2; 117/3; 184/2; 226/3

  Query Match             14.1%;  Score 233.5;  DB 2;  Length 262;
  Best Local Similarity   27.7%;  Pred. No. 1.6e-13;
  Matches   67;  Conservative   46;  Mismatches  100;  Indels   29;  Gaps    6;

Qy     43 NFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFFLLGT 99
          ::| ||: :       ::::| |     | |  :||||       :|   ::|  ||:|
Db     22 SWLYLPEIEETSQEKDILIIVASRTDSYARRNILRQTWMSKANSEIVANGRMKPLFLVGL 81

Qy    100 TSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSD 158
          |    : |: | ||:: :|||| |  | | |  || |::  : :     |:   : | | |
Db     82 TPGEYKMKKMVMQEAKLYGDIIVVDMNDNYEELTYKSLAILLYGVSKAPRYQMIGKIDED 141

Qy    159 MFINVDYLTELLLKKNRTTRFFTGFLKLNEFPI-----------RQPFSKWFVSKSEYP 206
          :    | ||||       :  ||:      |             |    :|:| :| |
Db    142 VMFFPDKLTEL---------YDQGFIDATPLRIYGLKMQSGANIFRDKTHRWYVPESSYS 192

Qy    207 WDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEELHSQPT 265
            ::| : ||  |: : : | |:   :|   :|:::|||: |:  | | | :: |   |
Db    193 CSKFPEYVSGMLYMVTWEAAQQIIKSTKYRDFIQVEDVFLTGILAEDLGISVKNL---PE 249

Qy    266 FF 267
          |:
Db    250 FY 251


RESULT 6
T20446
hypothetical protein E03H4.11 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T20446
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19276
A;Accession: T20446
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-384 <WIL>
A;Cross-references: EMBL:Z81492; PIDN:CAB04032.1; GSPDB:GN00019; CESP:E03H4.11
A;Experimental source: clone E03H4
C;Genetics:
A;Gene: CESP:E03H4.11
A;Map position: 1
A;Introns: 46/1; 76/1; 196/3; 263/2; 305/3; 335/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             13.5%;  Score 223.5;  DB 2;  Length 384;
  Best Local Similarity   27.0%;  Pred. No. 2.2e-12;
  Matches   67;  Conservative   44;  Mismatches  104;  Indels   33;  Gaps    6;

Qy     39 KKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFF 95
          :|:  :| ||: :       ::::| |     | |  :||||       :|   ::|  |
Db     97 QKNYTWLYLPEIEEASQEKDILMIVASRTDSYARRNIMRQTWMNKSDSEIVANGRMKPLF 156

Qy     96 LLGTTSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMK 154
          |:| |    : |: | ||:: :|||| |  | | |  || |::  : :     |:   : |
Db    157 LVGLTPGDYKMKKMVMQEAKLYGDIIVVDMNDTYEELTYKSLAILLYGVSKAPRYQMIGK 216

Qy    155 TDSDMFINVDYLTELL-------------LKKNRTTRFFTGFLKLNEFPIRQPFSKWFV 200
          | |:    | || |                |     | |           |   :|:|
Db    217 IDEDVIFFPDKLTALYEQGIIDATPLCAYGYKIQAGARIF----------RDKNDRWYV 265

Qy    201 SKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEE 259
           :| |   ::| : ||  |: : : | |:   :|   :|::||||: |:  | | | :
Db    266 PESSYSCSKFPEYVSGMLYMVTWEAAQQIIKSTKYRDFIQVEDVFLTGILAEDLGISVRN 325

Qy    260 LHSQPTFF 267
          |   | |:
Db    326 L---PKFY 330


RESULT 7
T18639
hypothetical protein B0024.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18639
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19001
A;Accession: T18639
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-507 <WIL>
A;Cross-references: EMBL:Z71178; PIDN:CAA94876.1; GSPDB:GN00023; CESP:B0024.3
A;Experimental source: clone B0024
C;Genetics:
A;Gene: CESP:B0024.3
A;Map position: 5
A;Introns: 69/2; 118/2; 143/2; 181/1; 246/2; 281/3; 393/2; 437/3

  Query Match             13.2%;  Score 219.5;  DB 2;  Length 507;
  Best Local Similarity   25.6%;  Pred. No. 7.2e-12;
  Matches   74;  Conservative   56;  Mismatches  138;  Indels   21;  Gaps    9;

Qy     31 FKEQSFVYKKDGNFLKLPDTDC-RQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMV 86
          ||:|:  |:    |: :|   |   |   : :| |: :  :  | |||:||       |
Db    227 FKDQNHTYQ----FITVPKKQCSNNTKLQITILSTAGNFDI--RQAIRETWANPNNSEHV 280

Qy     87 KGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFC 146
             ::  |::  ||:      : :| ::  |:|  |  : |  | ||    : :    |
Db    281 ANNDVRISFIISKTSNEFLNFALQKEIEKFDDMIVTDLYESYELLILKVHAILSYKQSHC 340

Qy    147 PQAAFVMKTDSDMFINVDYLTELLL-KKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEY 205
          | | :| | || :::| |   |  ||  :    :| :  |  |:|:   :|:| |: |
Db    341 QLADFQLKIDDDMAVDMDGLYRSLEDKKQASINGISGIIWKNSPPVREKKHRWYVPKTLY 400

Qy    206 PWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVF----VG--LCLERLNIRLEE 259
             :||:  |  |:  :  ::  :|:     :||||    :|  | ::::|
Db    401 SEKFFPPYIDGPIYLIGKNAVPRMLEEAKNYNQWIIEDVFWTGVIGKALKIKQINWANHL 460

Qy    260 LHSQPTFFPGGLRFSVCLFRRIVACHFIK-PRTLLDYWQALENSRGEDC 307
          |       |  |: |      | | | :| |: : | :| |   :|  |
Db    461 LRYVIELIPSRLKCSKGGVPLIYAVHNMKGPQMIHDGYQKL---KGVKC 506


RESULT 8
T20031
hypothetical protein C47F8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T20031; T22806
R;McLay, K.
```

```
submitted to the EMBL Data Library, December 1997
A;Reference number: Z19212
A;Accession: T20031
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-345 <WIL>
A;Cross-references: EMBL:AL009246; PIDN:CAA15841.1; GSPDB:GN00019; CESP:C47F8.3
A;Experimental source: clone C47F8
R;Kershaw, J.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19617
A;Accession: T22806
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-345 <WI2>
A;Cross-references: EMBL:Z81553; PIDN:CAB04502.1; GSPDB:GN00019; CESP:C47F8.3
A;Experimental source: clone F56H6
C;Genetics:
A;Gene: CESP:C47F8.3
A;Map position: 1
A;Introns: 23/1; 49/1; 169/3; 236/2; 278/3; 308/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             12.6%;  Score 209.5;  DB 2;  Length 345;
  Best Local Similarity   24.6%;  Pred. No. 3.5e-11;
  Matches   73;  Conservative   46;  Mismatches  129;  Indels   49;  Gaps    9;

Qy      11 ICLLVLGALCLYFSMYSLNPFKEQSFVYK---------------------------- 39
           |||:|:      |  : | |  : ||| :
Db      16 ICLIVV------FLFWKLQPNPKSSFVLEEHCVQSGWNISTLSPRNIDFGSSFIVSFANI 69

Qy      40 -KDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW-GKE--RMVKGKQLKTFF 95
            |   :: ||          :::|| |  |  | |  :| ||  ||   |:|  ::   |
Db      70 HKSQKWIYLPKMKNVTWKHDILMLVVSKTKNFARRNVLRSTWMNKENSEMMKSGRMHALF 129

Qy      96 LLGTTSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMK 154
             :|        |: | :|:: |||::  |  | | ||  ||:  : :      |   : |
Db     130 FVGLVPGDQNLKKLVLEEAEIHGDMVVVDLEDTYDNLPFKTLALLLYGTSKASQFKIIGK 189

Qy     155 TDSDMFINVDYLTELLLKK--NRTTRFFTGFLK-LNEFPIRQPFSKWFVSKSEYPWDRYP 211
             | |:    | |  :| :   |  |    | |    |  :|     |:| :: |    ||
Db     190 IDDDVMFFPDQLLPMLDRNFVNSNTLSIYGHLSTAEELVLRNKTEPWYVPETAYNCTVYP 249

Qy     212 PFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEELHSQPTFF 267
           :  |  |: : | || : : :    :: :||  : |:  ::| ||    :| |  |
Db     250 VYVMGPIYLVTKDAASLILDNANHQQFMTVEDALITGIIAQKLGIR---RYSLPNVF 303


RESULT 9
T20029
hypothetical protein C47F8.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T20029
R;McLay, K.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z19212
A;Accession: T20029
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <WIL>
A;Cross-references: EMBL:AL009246; PIDN:CAA15839.1; GSPDB:GN00019; CESP:C47F8.6
A;Experimental source: clone C47F8
C;Genetics:
A;Gene: CESP:C47F8.6
A;Map position: 1
A;Introns: 19/1; 52/1; 172/3; 239/2; 281/3; 311/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             12.1%;  Score 200;  DB 2;  Length 357;
  Best Local Similarity   25.7%;  Pred. No. 2.7e-10;
  Matches   59;  Conservative   48;  Mismatches  115;  Indels    8;  Gaps    4;

Qy      39 KKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFF 95
           :|:  :| ||  :        ::::| |      | |  :||||        :|   ::|: |
Db      73 QKNYTWLYLPTIEETSKETDILMIVASRTDSYARRNIMRQTWMNKSNSEIVANGRMKSLF 132

Qy      96 LLGTTSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMK 154
           |:|   :  : |: | ||:: :|||| |  | | |  |::|   :     ||   : |
Db     133 LVGLAPADYKVKKMVMQEAKLYGDIIIVDMDDTYEELIYKSLMIFLFGVSKAPQYKIIGK 192

Qy     155 TDSDMFINVDYLTELL---LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYP 211
             | |:    | |  |    :  :      :   :        |   ::|:| :| |   :||
Db     193 IDEDIMFFPDKLMALYEQGIIDSTPVSLYGLVIPAGRDIFRDKTNRWYVPESAYSCSQYP 252

Qy     212 PFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEEL 260
           : ||  |: : : |  :   :|   :|::||| : |:  | | | : ::
Db     253 AYLSGMYYMATREAAQMLLKSTKHRDFIQVEDVLLTGILAEDLGIPIRDM 302


RESULT 10
T24925
hypothetical protein T15D6.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T24925
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19956
A;Accession: T24925
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-376 <WIL>
A;Cross-references: EMBL:Z83125; PIDN:CAB05616.1; GSPDB:GN00019; CESP:T15D6.5
A;Experimental source: clone T15D6
C;Genetics:
A;Gene: CESP:T15D6.5
A;Map position: 1
A;Introns: 59/1; 87/1; 207/3; 274/2; 316/3; 339/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match             12.0%;  Score 199;  DB 2;  Length 376;
  Best Local Similarity   23.5%;  Pred. No. 3.5e-10;
  Matches   64;  Conservative   60;  Mismatches  120;  Indels   28;  Gaps   10;

Qy      39 KKDGNFLKLPDTDCRQTPPF---LVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLK 92
           :|   :| ||      : | :   ::::| |    :: |  :|:||       ::::  :::
Db     108 QKSQRWLHLPKI---KNPTWNRDILMIVASRPGSVSRRKVLRKTWMNKANSKIIRNGRMQ 164

Qy      93 TFFLLGTTSSAAE-TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAF 151
             ||:|  :   :  | | :|::  ||||  :  | | ||  | :  : :
Db     165 VLFLVGMVAGDRDLMKAVKKEAESFGDIIVMNLEDTYDNLPFKVLSLLLYGTNKASDFKI 224

Qy     152 VMKTDSDMFINVDYLTELLLKK--NRTTRFFTGFLKL-NEFPIRQPFSKWFVSKSEYPWD 208
           : | | |:    | || || :   : ::    |:|   :|  :|     |:| :: |
Db     225 IGKIDDDVIFFPDRLTPLLDENVIDSSSYSIYGYLSQDDELVVRNETKPWYVPETAYNCT 284

Qy     209 RYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFV-GLCLERLNIRLEELHSQPTFF 267
           :|| :  |  |: :   |: :   |:   :: :||  : |:  | | |   : || |
Db     285 KYPVYALGPFYLITNKAANLIVENSRFQNFMTVEDALIAGIIAEGLGI---QRHSLPMNT 341

Qy     268 PGGLRFSVCLFRRIVACHFIK--PRTLLDYWQ 297
           |         ::|:: |  |   |  ||::|
Db     342 DG---------KKILSWHMSKRSDRQFLDFYQ 364


RESULT 11
T20028
hypothetical protein C47F8.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T20028
R;McLay, K.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z19212
A;Accession: T20028
```

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-332 <WIL>
A;Cross-references: EMBL:AL009246; PIDN:CAA15838.1; GSPDB:GN00019; CESP:C47F8.5
A;Experimental source: clone C47F8
C;Genetics:
A;Gene: CESP:C47F8.5
A;Map position: 1
A;Introns: 26/1; 146/3; 213/2; 255/3; 285/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match           11.6%;  Score 192;  DB 2;  Length 332;
  Best Local Similarity  25.7%;  Pred. No. 1.3e-09;
  Matches   59;  Conservative   45;  Mismatches  118;  Indels    8;  Gaps    4;

Qy      39 KKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTW---GKERMVKGKQLKTFF 95
           :|   :| ||:||     | ::::| |     | |  :|:||       :|:  ::|  |
Db      47 QKTYKWLYLPETDIVLRSPDILMMVASRTDSFARRNVLRKTWMNKNYSEIVRDGRMKALF 106

Qy      96 LLGTTSSAAETKE-VDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMK 154
           |:|  |     :  | :|:: :||::  |  | | :|  |::  : :     |:   : |
Db     107 LVGMVSEDYRVRRIVMEEAKLYGDMVVIDLEDTYDDLPFKSLSLLLYAVSKAPEFKVIGK 166

Qy     155 TDSDMFINVDYLTELLLKK---NRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYP 211
           | |:    | |  ||  |        |:   ||  |  |::  : |:|    |    ||
Db     167 IDEDVMFFPDKLIPLLDGKVIDPDAAAFYGQLLKEGEPVIKKKDAHWYVPDYAYNCTGYP 226

Qy     212 PFCSGTGYVFSGDVASQVYNVSKSVPYIKLED-VFVGLCLERLNIRLEEL 260
           : :|  |: :   |  |   :|   :: :|| :  |:    | |  : |
Db     227 AYVAGPFYLATRKAAKLVLKFTKFQNFMTVEDSLITGILANDLGIPRKNL 276


RESULT 12
T20876
hypothetical protein F14B6.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C;Accession: T20876
R;White, S.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19338
A;Accession: T20876
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-376 <WIL>
A;Cross-references: EMBL:Z81502; PIDN:CAB04107.1; GSPDB:GN00019; CESP:F14B6.4
A;Experimental source: clone F14B6
C;Genetics:
A;Gene: CESP:F14B6.4
A;Map position: 1
A;Introns: 31/1; 69/1; 189/3; 256/2; 298/3; 328/3
C;Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

  Query Match           11.4%;  Score 189.5;  DB 2;  Length 376;
  Best Local Similarity  23.6%;  Pred. No. 2.6e-09;
  Matches   68;  Conservative   57;  Mismatches  138;  Indels   25;  Gaps   10;

Qy      39 KKDGNFLKLPDTDCRQT------PPF-------LVLLVTSSHKQLAERMAIRQTW---GK 82
           |  |:   :   | ::|      | |       ::::| |     | |  :|:||
Db      76 KDFGSSFTISFADIQKTHSWLYLPKFQNSMSSEILMIVASRTDSFARRNVLRKTWMNPEN 135

Qy      83 ERMVKGKQLKTFFLLGTT--SSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIE 140
           ::|  ::|  ||:| |    :   | | :|:: :|||::  |  | |  |  |::  :
Db     136 SEIIKDGRMKALFLVGMTDGDDSRMRKVVMEEARIYGDMVVVDLKDTYEELPFKSLTTLL 195

Qy     141 WVHRFCPQAAFVMKTDSDMFINVDYLTELLLKK--NRTTRFFTGFL-KLNEFPIRQPFSK 197
           :      :   : | | |:    | :  || :   : ::    | |     :  |    :
Db     196 YGTSKASEFKLIGKIDEDIMFFPDKILPLLEQNLIDPSSESIYGMLFAEGGYVYRDKEHR 255

Qy     198 WFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRL 257
           |||  | |  | :||:  |  |: : | | :: | :|   :| :||  :   |   : ::
Db     256 WFVPDSTYGCDMFPPYTGGLFYLVTQDAAKKILNATKHRIFIPIEDALINGILAN-DCKI 314

Qy     258 EELHSQPTFFPGGLRFSVCLFRRIVACHFIK-PRTLLDY-WQALENSR 303
           :|  |  :      :    | ::| |  |  :  ||| :: |  ||
Db     315 PRIH-LPEIYGDHHSLTDEDTRGVLAWHTGKNDQQFLDYFYRKLSESR 361


RESULT 13
B86353
protein F2E2.6 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: B86353
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alo
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B86353
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <STO>
A;Cross-references: GB:AE005172; NID:g9280694; PIDN:AAF86563.1; GSPDB:GN00141
C;Genetics:
A;Gene: F2E2.6
A;Map position: 1

  Query Match           10.4%;  Score 171.5;  DB 2;  Length 398;
  Best Local Similarity  26.0%;  Pred. No. 1.2e-07;
  Matches   59;  Conservative   45;  Mismatches   80;  Indels   43;  Gaps   11;

Qy      58 FLVLLVTSSHKQLAERMAIRQTW---GK--ERMVKGKQLKTFFLLG--TTSSAAETKEVD 110
           |:|: : ::      | ::|:||   |:  |:: | | :   |::|  :| ::   ||:|
Db     127 FMVIGINTAFSSRKRRDSLRETWMPQGEKLEKLEKEKGIVVKFMIGHSSTPNSMLDKEID 186

Qy     111 QESQRHGDIIQKDFLDVYYNLTLKTMMGI-----EWVHRFCPQAAFVMKTDSDMFINVDY 165
           |  :: |  : | :: ||||: ||         :|       | | :| | |: :|:
Db     187 SEDAQYNDFFRLDHVEGYYNLSAKTKSFFSSAVAKW------DAEFYVKIDDDVHVNLGT 240

Qy     166 LTELLLKKNRTTRFFTGFLKLNEFPI--------RQP-FSKWFVSKSEYPWDRYPPFCSG 216
           |   |       | : | :|    |:         |:| | |:    ::|       |
Db     241 LASTLASHRSKPRVYIGCMKSG--PVLTKKTAKYREPEFWKFGEEGNKY-------FRHA 291

Qy     217 TG--YVFSGDVASQVYNVSKSVPYIKLEDV-----FVGLCLERLNIR 256
           ||  |  | |:|: : |    :     |||     |:|| :|::: |
Db     292 TGQIYAISKDLATYISNNQPILHKYANEDVTLGSWFIGLEVEQIDDR 338


RESULT 14
C96573
protein F12M16.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96573
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alo
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C96573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-353 <STO>
```

```
A;Cross-references: GB:AE005173; NID:g7769857; PIDN:AAF69535.1; GSPDB:GN00141
C;Genetics:
A;Gene: F12M16.19
A;Map position: 1

  Query Match             9.7%;  Score 160;  DB 2;  Length 353;
  Best Local Similarity   23.6%;  Pred. No. 1.1e-06;
  Matches   61;  Conservative   42;  Mismatches  112;  Indels   44;  Gaps    8;

Qy      45 LKLPDTDC-RQTPPFLVLLV--------------TSSHKQL------------AERMAIR 77
           |:|   :|   |||  | :|                || :              | ::|
Db      47 LRLTGRNCLTNTPPKTVRVVWDVAGNSNGVVSGEKKRHKVMGFVGIQTGFGSAGRRRSLR 106

Qy      78 QTW------GKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNL 131
           :||      |  |: :   |   |::| | |  :  :: :|   : | :  |  : |  |
Db     107 KTWMPSDPEGLRRLEESTGLAIRFMIGKTKSEEKMAQLRREIAEYDDFVLLDIEEEYSKL 166

Qy     132 TLKTMMGIEWVHRFCPQAA-----FVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKL 186
             ||:: :  :  |    |     | :| | |:::  | |: || |:   :: : | ||
Db     167 PYKTLVRVICLAFFKAAYALYDSEFYVKADDDIYLRPDRLSLLLAKERSHSQTYLGCLK- 225

Qy     187 NEFPI-RQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVY---NVSKSVPYIKLE 242
            : |:   |  ||:   |      |     | | | || : :    |   ::     |
Db     226 -KGPVFTDPKLKWYEPLSHLLGKEYFLHAYGPIYALSADVVASLVALKNNRHALLMFNNE 284

Qy     243 DVFVGLCLERLNIRLEELH 261
           || :|  :  :|:  |  |
Db     285 DVTIGAWMLAMNVNHENHH 303


RESULT 15
T22387
hypothetical protein F48F7.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000`
C;Accession: T22387
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19559
A;Accession: T22387
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-253 <WIL>
A;Cross-references: EMBL:Z69661; PIDN:CAA93493.1; GSPDB:GN00028; CESP:F48F7.3
A;Experimental source: clone F48F7
C;Genetics:
A;Gene: CESP:F48F7.3
A;Map position: X
A;Introns: 13/3; 55/3; 93/2; 161/3; 219/2

  Query Match             9.6%;  Score 158.5;  DB 2;  Length 253;
  Best Local Similarity   23.2%;  Pred. No. 9.9e-07;
  Matches   48;  Conservative   45;  Mismatches   87;  Indels   27;  Gaps    5;

Qy      94 FFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVM 153
           :|::||        ::| | : : |||| |  : |:|:| | :  |  :::       ::
Db      19 YFIVGTDIDV----DIDDEIKTYEDIIQVDINENYHNITYKAIFWINEINKCRHGPKLLL 74

Qy     154 KTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 213
           | | |: |::  |  |: : :    |    :  |   :|   |||::|| :| :     :
Db      75 KIDDDVHIDMIGLQFLIKRYSVMDDFMACRVISNGQVVRNNTSKWYLSKKDYKFPNLGTY 134

Qy     214 CSGTGYVFSGDVASQVY-NVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLR 272
           | |  |  ||::   :: |: ||  :: ::| :|   |                |  :
Db     135 CQGMVYFISGNLLPILHGNIEKS-QFLWMDDWYVTRSL----------------VGDYK 176

Qy     273 FSVCLFRRIVACHFIKPRTLLDYWQAL 299
           |     :    | | | |: : : :|
Db     177 ISYYSLEQ----HSISPNTVREVYSSL 199


Search completed: April 14, 2003, 12:11:20
Job time : 48 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:41:40 ; Search time 33 Seconds
(without alignments)
1748.279 Million cell updates/sec

Title: US-09-914-152-1_COPY_31_310
Perfect score: 1498
Sequence: 1 FKEQSFVYKKDGNFLKLPDT.....TLLDYWQALENSRGEDCPPV 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1389.5	92.8	302	6	Q9N292	Q9n292 pongo pygma
2	577	38.5	326	4	Q9Y5Z6	Q9y5z6 homo sapien
3	577	38.5	326	6	Q9MYM7	Q9mym7 pongo pygma
4	577	38.5	326	11	O54904	O54904 mus musculu
5	572	38.2	305	11	Q920V5	Q920v5 mus spicile
6	572	38.2	305	11	Q91V52	Q91v52 mus musculu
7	569	38.0	422	4	O43825	O43825 homo sapien
8	567	37.9	409	11	Q920V4	Q920v4 mus musculu
9	567	37.9	409	11	Q920V3	Q920v3 mus musculu
10	567	37.9	409	11	Q920V2	Q920v2 mus spicile
11	567	37.9	409	11	Q91VE9	Q91ve9 mus musculu
12	567	37.9	409	11	Q91V58	Q91v58 mus musculu
13	567	37.9	409	11	Q91V19	Q91v19 mus musculu
14	567	37.9	422	11	O54905	O54905 mus musculu
15	565	37.7	422	4	Q9BZQ9	Q9bzq9 homo sapien
16	555.5	37.1	331	11	Q920V1	Q920v1 mus musculu
17	555.5	37.1	331	11	Q920V0	Q920v0 mus musculu
18	555.5	37.1	331	11	Q91V72	Q91v72 mus musculu
19	555.5	37.1	331	11	O54906	O54906 mus spicile
20	548.5	36.6	331	4	O75752	O75752 homo sapien
21	529.5	35.3	255	11	Q9CTE5	Q9cte5 mus musculu
22	433.5	28.9	412	13	Q8UWM2	Q8uwm2 brachydanio
23	429	28.6	204	4	Q8TDY1	Q8tdy1 homo sapien
24	413	27.6	372	4	Q9C0J2	Q9c0j2 homo sapien
25	406	27.1	390	13	Q8UWM1	Q8uwm1 brachydanio
26	402.5	26.9	372	11	Q9D7Z2	Q9d7z2 mus musculu
27	401.5	26.8	372	11	Q8R0U2	Q8r0u2 mus musculu
28	396	26.4	378	4	Q9BYG0	Q9byg0 homo sapien
29	393	26.2	371	11	O88178	O88178 rattus norv
30	388	25.9	370	11	Q920U7	Q920u7 mus spicile
31	388	25.9	370	11	Q91VC1	Q91vc1 mus musculu
32	388	25.9	371	11	Q9Z0F0	Q9z0f0 mus musculu
33	388	25.9	406	13	Q8UWM0	Q8uwm0 brachydanio
34	387.5	25.9	378	4	O96024	O96024 homo sapien
35	384	25.6	418	13	Q8UWM4	Q8uwm4 brachydanio
36	383	25.6	370	11	Q920U8	Q920u8 mus musculu
37	382.5	25.5	277	4	Q8TAZ4	Q8taz4 homo sapien
38	382.5	25.5	384	4	Q8TDX1	Q8tdx1 homo sapien
39	381.5	25.5	574	5	Q9VLS8	Q9vls8 drosophila
40	381.5	25.5	585	5	Q95TQ8	Q95tq8 drosophila
41	380	25.4	370	11	Q920U9	Q920u9 mus musculu
42	364	24.3	412	13	Q8UWM3	Q8uwm3 brachydanio
43	362	24.2	399	11	Q8VI16	Q8vi16 mus musculu
44	359.5	24.0	377	4	Q96EK0	Q96ek0 homo sapien
45	359	24.0	397	11	Q91V18	Q91v18 mus musculu

ALIGNMENTS

RESULT 1
Q9N292
ID Q9N292 PRELIMINARY; PRT; 302 AA.
AC Q9N292;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5 (Fragment).
GN BETA1,3-GALT 5.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORAN-PO17;
RA Liu Y., Saitou N.;
RT "Silver Project.";
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB041417; BAA94502.1; -.
DR InterPro; IPR002659; Galactosyl_T.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
FT NON_TER 302 302
SQ SEQUENCE 302 AA; 35500 MW; 51A555C864465A90 CRC64;

Query Match 92.8%; Score 1389.5; DB 6; Length 302;
Best Local Similarity 96.7%; Pred. No. 7.1e-126;
Matches 263; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

```
Qy       1 FKEQSFVYKK-DGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGK 59
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||| ||||
Db      31 FKEQSFVYKKEDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGK 90

Qy      60 QLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQA 119
           |||||||||||||||| ||||||||||||||||||||||||||||:||||||||||
Db      91 QLKTFFLLGTTSSAAEMKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGMEWVHRFCPQA 150

Qy     120 AFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDR 179
```

```
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    151 AFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNELPIRQPFSKWFVSKSEYPWDR 210

Qy    180 YPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPG 239
          |||||||| ||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db    211 YPPFCSGTAYVFSGDVASQVYNVSESVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPE 270

Qy    240 GLRFSVCLFRRIVACHFIKPRTLLDYWQALEN 271
          ||||||| ||||||||||||||||||||||||
Db    271 GLRFSVCRFRRIVACHFIKPRTLLDYWQALEN 302


RESULT 2
Q9Y5Z6
ID   Q9Y5Z6      PRELIMINARY;      PRT;   326 AA.
AC   Q9Y5Z6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE   1 (EC 2.4.1.-) (BETA3GALT1).
GN   B3GALT1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=COLON ADENOCARCINOMA;
RX   MEDLINE=99283878; PubMed=10356986;
RA   Bardoni A., Valli M., Trinchera M.;
RT   "Differential expression of beta1,3galactosyltransferases in human
RT   colon cells derived from adenocarcinomas or normal mucosa.";
RL   FEBS Lett. 451:75-80(1999).
RN   [2]
RP   FUNCTION, AND TISSUE SPECIFICITY.
RX   MEDLINE=98250717; PubMed=9582303;
RA   Amado M., Almeida R., Carneiro F., Levery S.B., Holmes E.H.,
RA   Nomoto M., Hollingsworth M.A., Hassan H., Schwientek T., Nielsen P.A.,
RA   Bennett E.P., Clausen H.;
RT   "A family of human beta3-galactosyltransferases. Characterization of
RT   four members of a UDP-galactose:beta-N-acetyl-glucosamine/beta-
RT   nacetyl-galactosamine beta-1,3-galactosyltransferase family.";
RL   J. Biol. Chem. 273:12770-12778(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Liu Y., Saitou N.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC       COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC       WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT1 AND GALT2
CC       ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH DIFFERENT
CC       AFFINITY.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC       GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC       GLUCOSAMINYLGLYCOPEPTIDE.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND COLON MUCOSA AND TO A
CC       LESSER EXTENT IN COLON ADENOCARCINOMA CELLS. NO EXPRESSION IN
CC       HEART, PLACENTA, KIDNEY, LIVER, SKELETAL MUSCLE, LUNG AND
CC       PANCREAS.
DR   EMBL; AF117222; AAD23451.1; -.
DR   EMBL; AB041407; BAA94492.1; -.
DR   MIM; 603093; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Glycoprotein; Golgi stack.
FT   DOMAIN        1      4       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      5     20       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       21    326       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   326 AA;  37993 MW;  83271E99B2EE74F5 CRC64;

  Query Match             38.5%;  Score 577;  DB 4;  Length 326;
  Best Local Similarity   42.8%;  Pred. No. 1.9e-47;
  Matches  116;  Conservative   46;  Mismatches   93;  Indels   16;  Gaps    4;

Qy     14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
          ||      | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db     65 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 124

Qy     74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
             : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||||:|:|:
Db    125 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 184

Qy    134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
          | |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
Db    185 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 243

Qy    192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
          | |||  :|  |     : ||||:||||| :| |           |   |     : :|:|
Db    244 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 295

Qy    247 LFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
            :||::  | | |  :   |  : : :   |
Db    296 RYRRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 3
Q9MYM7
ID   Q9MYM7      PRELIMINARY;      PRT;   326 AA.
AC   Q9MYM7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Beta 1,3-galactosyltransferase polypeptide 1.
GN   B3GALT1.
OS   Pongo pygmaeus (Orangutan),
OS   Pan troglodytes (Chimpanzee),
OS   Pan paniscus (Pygmy chimpanzee) (Bonobo), and
OS   Gorilla gorilla (gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600, 9598, 9597, 9593;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=P.pygmaeus, P.troglodytes, P.paniscus, and G.gorilla;
RA   Liu Y., Saitou N.;
RT   "Silver Project.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB041411; BAA94496.1; -.
DR   EMBL; AB041408; BAA94493.1; -.
DR   EMBL; AB041409; BAA94494.1; -.
DR   EMBL; AB041410; BAA94495.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   326 AA;  37993 MW;  83271E99B2EE74F5 CRC64;

  Query Match             38.5%;  Score 577;  DB 6;  Length 326;
  Best Local Similarity   42.8%;  Pred. No. 1.9e-47;
  Matches  116;  Conservative   46;  Mismatches   93;  Indels   16;  Gaps    4;

Qy     14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
          ||      | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db     65 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 124

Qy     74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
             : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||||:|:|:
Db    125 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 184

Qy    134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
          | |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
```

```
Db    185 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 243

Qy    192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
          | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db    244 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 295

Qy    247 LFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
           :||::  | | |  :   |  : : :   |
Db    296 RYRRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 4
O54904
ID   O54904      PRELIMINARY;      PRT;   326 AA.
AC   O54904;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE   II (EC 2.4.1.-) (UDP-GAL:BETAGLCNAC BETA 1,3-GALACTOSYLTRANFERASE-I)
DE   (BETA3GALT-I).
GN   B3GT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=129/SVJ;
RX   MEDLINE=98079027; PubMed=9417047;
RA   Hennet T., Dinter A., Kuhnert P., Mattu T.S., Rudd P.M., Berger E.G.;
RT   "Genomic cloning and expression of three murine UDP-galactose: beta-N-
RT   acetylglucosamine beta1,3-galactosyltransferase genes.";
RL   J. Biol. Chem. 273:58-65(1998).
CC   -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC       COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC       WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT-I, GALT-II
CC       AND GALT-III ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH
CC       DIFFERENT AFFINITY.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC       GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC       GLUCOSAMINYLGLYCOPEPTIDE.
CC   -!- COFACTOR: MANGANESE.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND TO A LESSER EXTENT IN
CC       HEART, KIDNEY, LIVER, LUNG, OVARY, TESTIS, UTERUS, COLON, STOMACH,
CC       BONE MARROW, THYMUS, SPLEEN AND LYMPH NODES.
DR   EMBL; AF029790; AAC53523.1; -.
DR   MGD; MGI:1349403; B3galt1.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Glycoprotein; Golgi stack.
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     25       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       26    326       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   326 AA;  37979 MW;  9825D68869EFAD3D CRC64;

  Query Match             38.5%;  Score 577;  DB 11;  Length 326;
  Best Local Similarity   42.8%;  Pred. No. 1.9e-47;
  Matches  116;  Conservative   46;  Mismatches   93;  Indels   16;  Gaps    4;

Qy     14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
          ||      | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db     65 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 124

Qy     74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
             : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||:|:|:
Db    125 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 184

Qy    134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
          | |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
Db    185 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 243

Qy    192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
          | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db    244 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 295

Qy    247 LFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
           :||::  | | |  :   |  : : :   |
Db    296 RYRRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 5
Q920V5
ID   Q920V5      PRELIMINARY;      PRT;   305 AA.
AC   Q920V5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase I (Fragment).
GN   B3GT1.
OS   Mus spicilegus (Steppe mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZBN;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039143; BAB68667.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     305    305
SQ   SEQUENCE   305 AA;  35611 MW;  A338019F3B06EF5A CRC64;

  Query Match             38.2%;  Score 572;  DB 11;  Length 305;
  Best Local Similarity   44.2%;  Pred. No. 5.4e-47;
  Matches  115;  Conservative   43;  Mismatches   86;  Indels   16;  Gaps    4;

Qy     14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
          ||      | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db     54 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 113

Qy     74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
             : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||:|:|:
Db    114 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 173

Qy    134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
          | |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
Db    174 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 232

Qy    192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
          | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db    233 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 284

Qy    247 LFRRIVACHFIKPRTLLDYW 266
           :||::  | | |  :   |
Db    285 RYRRVITVHQISPEEMHRIW 304


RESULT 6
Q91V52
ID   Q91V52      PRELIMINARY;      PRT;   305 AA.
AC   Q91V52;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
```

```
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase I (Fragment).
GN   B3GT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VARIOUS STRAINS;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039134; BAB68658.1; -.
DR   EMBL; AB039135; BAB68659.1; -.
DR   EMBL; AB039136; BAB68660.1; -.
DR   EMBL; AB039137; BAB68661.1; -.
DR   EMBL; AB039138; BAB68662.1; -.
DR   EMBL; AB039139; BAB68663.1; -.
DR   EMBL; AB039140; BAB68664.1; -.
DR   EMBL; AB039141; BAB68665.1; -.
DR   EMBL; AB039142; BAB68666.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1       1
FT   NON_TER     305     305
SQ   SEQUENCE   305 AA;  35611 MW;  A338019F3B06EF5A CRC64;

  Query Match             38.2%;  Score 572;  DB 11;  Length 305;
  Best Local Similarity   44.2%;  Pred. No. 5.4e-47;
  Matches  115;  Conservative   43;  Mismatches   86;  Indels   16;  Gaps    4;

Qy      14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
           ||      | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db      54 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 113

Qy      74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
             : |:||||   ||| :||:| |:||||||:||: ||  || :| :||||||:|:|:
Db     114 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 173

Qy     134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
           | |  .||| :     |:|||:: :|. |||   |||:: :  ||   ||||||||||:|
Db     174 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 232

Qy     192 SGDVASQVYYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
           | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db     233 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 284

Qy     247 LFRRIVACHFIKPRTLLDYW 266
            :||::  | | |  :  |
Db     285 RYRRVITVHQISPEEMHRIW 304


RESULT 7
O43825
ID   O43825      PRELIMINARY;      PRT;   422 AA.
AC   O43825;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE   2 (EC 2.4.1.-) (BETA3GALT2).
GN   GALT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   TISSUE=BRAIN;
RX   MEDLINE=98079080; PubMed=9417100;
RA   Kolbinger F., Streiff M.B., Katopodis A.G.;
RT   "Cloning of a human UDP-galactose:2-acetamido-2-deoxy-D-glucose 3beta-
RT   galactosyltransferase catalyzing the formation of type 1 chains.";
RL   J. Biol. Chem. 273:433-440(1998).
RN   [2]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE=98250717; PubMed=9582303;
RA   Amado M., Almeida R., Carneiro F., Levery S.B., Holmes E.H.,
RA   Nomoto M., Hollingsworth M.A., Hassan H., Schwientek T., Nielsen P.A.,
RA   Bennett E.P., Clausen H.;
RT   "A family of human beta3-galactosyltransferases. Characterization of
RT   four members of a UDP-galactose:beta-N-acetyl-glucosamine/beta-
RT   nacetyl-galactosamine beta-1,3-galactosyltransferase family.";
RL   J. Biol. Chem. 273:12770-12778(1998).
CC   -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC       COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC       WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT1 AND GALT2
CC       ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH DIFFERENT
CC       AFFINITY.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC       GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC       GLUCOSAMINYLGLYCOPEPTIDE.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN HEART AND TO A LESSER EXTENT IN
CC       BRAIN. NO EXPRESSION IN PLACENTA, KIDNEY, LIVER, SKELETAL MUSCLE,
CC       LUNG AND PANCREAS.
DR   EMBL; Y15014; CAA75245.1; -.
DR   EMBL; Y15060; CAA75344.1; -.
DR   MIM; 603018; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Glycoprotein; Golgi stack.
FT   DOMAIN        1      24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25      43       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       44     422       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     75      75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    100     100       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119     119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    176     176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226     226       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   422 AA;  49213 MW;  BA62942D6AFB4B0C CRC64;

  Query Match             38.0%;  Score 569;  DB 4;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 1.6e-46;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
             :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
           ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     258 KPRTLLDYWQALENSRGEDC 277
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405


RESULT 8
Q920V4
ID   Q920V4      PRELIMINARY;      PRT;   409 AA.
AC   Q920V4;
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CAST/EI;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039147; BAB68671.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47247 MW;  2425654F23A8F0E3 CRC64;

  Query Match             37.9%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-46;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
             :||::: |||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy     198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
           ::: ||  :  : ||||:||:|| :| :      ::  |   : :| | :  ::  |
Db     320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy     258 KPRTLLDYWQALENSRGEDC 277
           :|  |: ||  |: ::   |
Db     380 QPSELIKYWNHLQQNKHNAC 399


RESULT 9
Q920V3
ID   Q920V3      PRELIMINARY;      PRT;   409 AA.
AC   Q920V3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HMI/MSF;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039151; BAB68675.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47220 MW;  2D2EBC4D68E8463D CRC64;

  Query Match             37.9%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-46;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
             :||::: |||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy     198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
           ::: ||  :  : ||||:||:|| :| :      ::  |   : :| | :  ::  |
Db     320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy     258 KPRTLLDYWQALENSRGEDC 277
           :|  |: ||  |: ::   |
Db     380 QPSELIKYWNHLQQNKHNAC 399


RESULT 10
Q920V2
ID   Q920V2      PRELIMINARY;      PRT;   409 AA.
AC   Q920V2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus spicilegus (Steppe mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZBN;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039153; BAB68677.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47383 MW;  0DC51AB6335A58DE CRC64;

  Query Match             37.9%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-46;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
             :||::: |||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319
```

```
Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db    320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    380 QPSELIKYWNHLQQNKHNAC 399


RESULT 11
Q91VE9
ID   Q91VE9      PRELIMINARY;      PRT;   409 AA.
AC   Q91VE9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BLG2/MSF, AND NJL/MSF;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039146; BAB68670.1; -.
DR   EMBL; AB039152; BAB68676.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47196 MW;  06D262B08B0B90F4 CRC64;

  Query Match             37.9%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-46;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy     22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
          |::  |||:||: :   |:  | |||||||.| :  | |:   |||| :   :      :
Db    140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy     80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
           :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db    200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy    140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
          ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db    260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db    320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    380 QPSELIKYWNHLQQNKHNAC 399


RESULT 12
Q91V58
ID   Q91V58      PRELIMINARY;      PRT;   409 AA.
AC   Q91V58;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/10SNJ, BFM/2MSF, AND PGN2;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039144; BAB68668.1; -.
DR   EMBL; AB039145; BAB68669.1; -.
DR   EMBL; AB039149; BAB68673.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47233 MW;  9DBE67873273F13A CRC64;

  Query Match             37.9%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-46;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy     22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
          |::  |||:||: :   |:  | |||||||.| :  | |:   |||| :   :      :
Db    140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy     80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
           :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db    200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy    140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
          ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db    260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db    320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    380 QPSELIKYWNHLQQNKHNAC 399


RESULT 13
Q91V19
ID   Q91V19      PRELIMINARY;      PRT;   409 AA.
AC   Q91V19;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSM/MSF, AND SWN/MSF;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039148; BAB68672.1; -.
DR   EMBL; AB039150; BAB68674.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47206 MW;  D7BD514EE40A0157 CRC64;
```

```
  Query Match             37.9%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-46;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
            :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy     198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
           ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db     320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy     258 KPRTLLDYWQALENSRGEDC 277
           :|  |: ||  |: ::   |
Db     380 QPSELIKYWNHLQQNKHNAC 399


RESULT 14
O54905
ID   O54905      PRELIMINARY;      PRT;   422 AA.
AC   O54905;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE   II (EC 2.4.1.-) (UDP-GAL:BETAGLCNAC BETA 1,3-GALACTOSYLTRANFERASE-II)
DE   (BETA3GALT-II).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=129/SVJ;
RX   MEDLINE=98079027; PubMed=9417047;
RA   Hennet T., Dinter A., Kuhnert P., Mattu T.S., Rudd P.M., Berger E.G.;
RT   "Genomic cloning and expression of three murine UDP-galactose: beta-N-
RT   acetylglucosamine beta1,3-galactosyltransferase genes.";
RL   J. Biol. Chem. 273:58-65(1998).
CC   -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC       COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC       WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT-I, GALT-II
CC       AND GALT-III ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH
CC       DIFFERENT AFFINITY.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC       GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC       GLUCOSAMINYLGLYCOPEPTIDE.
CC   -!- COFACTOR: MANGANESE.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND HEART. BARELY DETECTED
CC       IN OVARY, COLON AND LYMPH NODES. NO EXPRESSED IN TESTIS, UTERUS,
CC       STOMACH, KIDNEY, LIVER, LUNG, BONE MARROW, THYMUS AND SPLEEN.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-14 IS THE INITIATOR.
DR   EMBL; AF029791; AAC53524.1; -.
DR   MGD; MGI:1349461; B3galt2.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Glycoprotein; Golgi stack; Manganese.
FT   DOMAIN        1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25     43       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       44    422       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   422 AA;  49107 MW;  D03D18A7484DCB03 CRC64;

  Query Match             37.9%;  Score 567;  DB 11;  Length 422;
  Best Local Similarity   41.2%;  Pred. No. 2.5e-46;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :      :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 205

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
            :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db     206 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 265

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
           ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     258 KPRTLLDYWQALENSRGEDC 277
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405


RESULT 15
Q9BZQ9
ID   Q9BZQ9      PRELIMINARY;      PRT;   422 AA.
AC   Q9BZQ9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   B3GALT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21218927; PubMed=11318611;
RA   Sood R., Bonner T.I., Malakowska I., Stephan D.A., Robbins C.M.,
RA   Connors T.D., Morgenbesser S.D., Su K., Faruque M.U., Pinkett H.,
RA   Graham C., Baxevanis A.D., Klinger K.W., Landes G.M., Trent J.M.,
RA   Carpten J.D.;
RT   "Cloning and characterization of 13 novel transcripts and the human
RT   RGS8 gene from the 1q25 region encompassing the hereditary prostate
RT   cancer (hpc1) locus.";
RL   Genomics 73:211-222(2001).
DR   EMBL; AF288390; AAG60610.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
SQ   SEQUENCE   422 AA;  49241 MW;  BA6294E26AFB4ADC CRC64;

  Query Match             37.7%;  Score 565;  DB 4;  Length 422;
  Best Local Similarity   41.2%;  Pred. No. 3.8e-46;
  Matches  107;  Conservative   54;  Mismatches   95;  Indels    4;  Gaps    2;

Qy      22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
           |::  |||:||: :   |:  | |||:||| | :  | |:   |||| :   :  :  : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRRTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy      80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
            :||::: ||||:::|| |||||:|||:||: ||  :||   :|||||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
```

```
          ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db    266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db    326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    386 QPSELIKYWNHLQQNKHNAC 405


Search completed: April  4, 2003, 14:45:05
Job time : 35 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:43:40 ; Search time 15 Seconds
(without alignments)
549.228 Million cell updates/sec

Title: US-09-914-152-1_COPY_31_310
Perfect score: 1498
Sequence: 1 FKEQSFVYKKDGNFLKLPDT.....TLLDYWQALENSRGEDCPPV 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	577	38.5	326	2	US-09-055-097-6	Sequence 6, Appli
2	387.5	25.9	378	2	US-09-055-097-1	Sequence 1, Appli
3	357	23.8	378	4	US-09-482-180A-2	Sequence 2, Appli
4	340	22.7	325	2	US-09-055-097-5	Sequence 5, Appli
5	314	21.0	397	4	US-09-459-133-2	Sequence 2, Appli
6	305	20.4	389	4	US-09-459-133-13	Sequence 13, Appl
7	138	9.2	372	1	US-08-207-904-10	Sequence 10, Appl
8	81.5	5.4	1513	5	PCT-US93-03076-2	Sequence 2, Appli
9	79.5	5.3	612	3	US-09-212-971-14	Sequence 14, Appl
10	79.5	5.3	612	4	US-08-800-929A-14	Sequence 14, Appl
11	79.5	5.3	612	4	US-08-569-749-14	Sequence 14, Appl
12	79.5	5.3	612	4	US-09-617-053A-14	Sequence 14, Appl
13	79.5	5.3	612	5	PCT-US96-12860-14	Sequence 14, Appl
14	78	5.2	438	5	PCT-US95-05922A-2	Sequence 2, Appli
15	78	5.2	618	4	US-08-569-749-2	Sequence 2, Appli
16	78	5.2	618	4	US-09-069-023-29	Sequence 29, Appl
17	78	5.2	618	5	PCT-US96-12860-2	Sequence 2, Appli
18	78	5.2	740	3	US-08-276-968A-18	Sequence 18, Appl
19	77	5.1	741	3	US-08-276-968A-20	Sequence 20, Appl
20	76	5.1	454	3	US-08-276-968A-22	Sequence 22, Appl
21	76	5.1	923	4	US-09-397-885-1	Sequence 1, Appli
22	75.5	5.0	331	4	US-09-134-001C-3626	Sequence 3626, Ap
23	75.5	5.0	821	1	US-09-377-465A-2	Sequence 2, Appli
24	75	5.0	740	1	US-08-016-863-16	Sequence 16, Appl
25	75	5.0	740	3	US-08-276-968A-16	Sequence 16, Appl
26	73.5	4.9	329	4	US-09-651-200-19	Sequence 19, Appl
27	73.5	4.9	340	4	US-09-651-200-2	Sequence 2, Appli
28	73.5	4.9	441	4	US-09-651-200-4	Sequence 4, Appli
29	73.5	4.9	534	4	US-09-651-200-6	Sequence 6, Appli
30	73.5	4.9	534	4	US-09-651-200-24	Sequence 24, Appl
31	73.5	4.9	802	4	US-09-433-043B-120	Sequence 120, App
32	73.5	4.9	2237	1	US-08-354-973-1	Sequence 1, Appli
33	73.5	4.9	3177	2	US-08-477-451-4	Sequence 4, Appli
34	72.5	4.8	289	1	US-07-593-657-7	Sequence 7, Appli
35	72.5	4.8	289	4	US-08-942-012B-4	Sequence 4, Appli
36	72.5	4.8	517	1	US-08-457-274A-2	Sequence 2, Appli
37	72.5	4.8	517	5	PCT-US95-05758-2	Sequence 2, Appli
38	72	4.8	363	3	US-08-586-165-7	Sequence 7, Appli
39	72	4.8	732	1	US-08-481-626-2	Sequence 2, Appli
40	72	4.8	732	4	US-08-989-299-4	Sequence 4, Appli
41	72	4.8	887	1	US-08-327-494A-4	Sequence 4, Appli
42	72	4.8	887	5	PCT-US95-13659-4	Sequence 4, Appli
43	72	4.8	1306	4	US-08-989-299-7	Sequence 7, Appli
44	71.5	4.8	417	1	US-08-351-981-7	Sequence 7, Appli
45	71.5	4.8	517	1	US-08-457-274A-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-055-097-6
; Sequence 6, Application US/09055097
; Patent No. 5955282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,097
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 2745735
US-09-055-097-6

Query Match 38.5%; Score 577; DB 2; Length 326;

```
  Best Local Similarity   42.8%;  Pred. No. 1.9e-61;
  Matches  116;  Conservative   46;  Mismatches   93;  Indels   16;  Gaps    4;

Qy      14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
           ||      | :  ||||:|:::::||:   | |||:||| |   || :: | ||||  :
Db      65 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 124

Qy      74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
             : |:||||   ||| :||:| |:|||||:||: ||  || :| :|||||||:|:|:
Db     125 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 184

Qy     134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
           | |   ||| :    |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
Db     185 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 243

Qy     192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
           | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db     244 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 295

Qy     247 LFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
           :||::  | | |  :   |  : : :   |
Db     296 RYRRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 2
US-09-055-097-1
; Sequence 1, Application US/09055097
; Patent No. 5955282
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  HUMAN OXIDIZED LDL RECEPTOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/055,097
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0490 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  DUODNOT02
;      CLONE:  1705085
US-09-055-097-1

  Query Match             25.9%;  Score 387.5;  DB 2;  Length 378;
  Best Local Similarity   30.2%;  Pred. No. 2.4e-38;
  Matches   88;  Conservative   57;  Mismatches  105;  Indels   41;  Gaps    6;

Qy      26 PPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSS-----AAETKEVD 80
           ||||::|| :: : | :| ||| :|| |  :| ::: |||| ::      ::  ::
Db      70 PPFLLILVCTAPENLNQRNAIRASWGGLREARGLRVQTLFLLGEPNAQHPVWGSQGSDLA 129

Qy      81 QESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV-DYLTEL 139
           ||   |||:|  | | | ||||||: |: |  : || | :|:||| |:::|| : ::||
Db     130 SESAAQGDILQAAFQDSYRNLTLKTLSGLNWAEKHCPMARYVLKTDDDVYVNVPELVSEL 189

Qy     140 LLKKNRTTRF-------------------------FTGFLKLNEFPIRQPFSKWFVSKS 173
           :|:  |  ::                         : | :     | | |  :  ||:
Db     190 VLRGGRWGQWERSTEPQREAEQEGGQVLHSEEVPLLYLGRVHWRVNPSRTPGGRHRVSEE 249

Qy     174 EYP--WDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELH 231
           ::|  |  :||: |||||| |       :  |:   | : |||||||:   |  :
Db     250 QWPHTWGPFPPYASGTGYVLSASAVQLILKVASRAPLLPLEDVFVGVSARRGGL----AP 305

Qy     232 SQPTFFPGGLRFSV---CLFRRIVACHFIKPRTLLDYWQALENSRGEDCPP 279
           :|    |   : :   |  : ::  | : |  : : |: :  | ||   |
Db     306 TQCVKLAGATHYPLDRCCYGKFLLTSHRLDPWKMQEAWKLVGGSDGERTAP 356


RESULT 3
US-09-482-180A-2
; Sequence 2, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Gao, Zeren
;  APPLICANT:  Whitmore, Theodore E.
;  APPLICANT:  Jaspers, Stephen
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
;  TITLE OF INVENTION:  BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZNSSP6
;  FILE REFERENCE: 98-80
;  CURRENT APPLICATION NUMBER: US/09/482,180A
;  CURRENT FILING DATE:  2000-01-12
;  PRIOR APPLICATION NUMBER: 60/115,721
;  PRIOR FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 378
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-482-180A-2

  Query Match             23.8%;  Score 357;  DB 4;  Length 378;
  Best Local Similarity   32.6%;  Pred. No. 1.2e-34;
  Matches   90;  Conservative   42;  Mismatches  134;  Indels   10;  Gaps    5;

Qy       8 YKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGK-ERMVKGKQLKTFFL 66
           |:   ||  | :         ||:| : |    :  | ||| |||:     :|:|||  ||
Db      99 YRHCRNFSILLEPSGCSKDTFLLLAIKSQPGHVERRAAIRSTWGRVGGWARGRQLKLVFL 158

Qy      67 LGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTD 126
           ||   ||   : :  ||:   ||:| || : ::||||| :    ||   |||| |::| |
Db     159 LGVAGSAPPAQLLAYESREFDDILQWDFTEDFFNLTLKELHLQRWVVAACPQAHFMLKGD 218

Qy     127 SDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSG 186
           |:|::|  : | |   :       | :     | |    |:|:  | |    |||:  |
Db     219 DDVFVHVPNVLEFLDGWDPAQDLLVGDVIRQALPNRNTKVKYFIPPSMYRATHYPPYAGG 278

Qy     187 TGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLR---- 242
           ||| |      ::  : :    : ::|||||:|| ||   |  :|    |   |:|
Db     279 GGYVMSRATVRRLQAIMEDAELLSIDDVFVGMCLRRLG--LSPMH-HAGFKTFGIRRPLD 335

Qy     243 -FSVCLFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
            ||:| ::  | : |  :   | ||    |  |
Db     336 PLDPCLYRGLLLVHRLSPLEMWTMW-ALVTDEGLKC 370
```

```
RESULT 4
US-09-055-097-5
; Sequence 5, Application US/09055097
; Patent No. 5955282
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  HUMAN OXIDIZED LDL RECEPTOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/055,097
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C.
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0490 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1150971
US-09-055-097-5

  Query Match             22.7%;  Score 340;  DB 2;  Length 325;
  Best Local Similarity   30.4%;  Pred. No. 1.1e-32;
  Matches   79;  Conservative   48;  Mismatches  121;  Indels   12;  Gaps    6;

Qy      14 FLKLPDTDCR---QTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT 70
           :|::|          |  | :|: |:       | |||:||| |        |:  |||||
Db      62 YLRVPSFTAEVPVDQPARLTMLIKSAVGNSRRREAIRRTWGYEGRFSDVHLRRVFLLGTA 121

Qy      71 SSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF 130
             :   |:|  ||: ||||:| || | |:| |||||:|: |       :: | :  | | :
Db     122 EDS--EKDVAWESREHGDILQADFTDAYFNNTLKTMLGMRWASEQFNRSEFYLFVDDDYY 179

Qy     131 INVDYLTELL---LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGT 187
           ::   : : |    : ::    | | :     |:|  ||||:|| |||:||:||: :
Db     180 VSAKNVLKFLGRGRQSHQPELLFAGHV-FQTSPLRHKFSKWYVSLEEYPFDRWPPYVTAG 238

Qy     188 GYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCL 247
           :: |     |:|  |  :|  : :||::|:   :  | |:  |      |
Db     239 AFILSQKALRQLYAASVHLPLFRFDDVYLGIVALKAGISLQ--HCDDFRFHRPAYKGPDS 296

Qy     248 FRRIVACH-FIKPRTLLDYW 266
           :  ::| | |  |  :   |
```

```
Db     297 YSSVIASHEFGDPEEMTRVW 316


RESULT 5
US-09-459-133-2
; Sequence 2, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Jaspers, Stephen R.
;  APPLICANT:  Gao, Zeren
;  TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
;  FILE REFERENCE: 98-77
;  CURRENT APPLICATION NUMBER: US/09/459,133
;  CURRENT FILING DATE:  1999-12-10
;  PRIOR APPLICATION NUMBER: 60/111,697
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 397
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (137)...(137)
;   OTHER INFORMATION: Xaa is Gly or Ser
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(397)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-459-133-2

  Query Match             21.0%;  Score 314;  DB 4;  Length 397;
  Best Local Similarity   29.9%;  Pred. No. 2.2e-29;
  Matches   78;  Conservative   50;  Mismatches  117;  Indels   16;  Gaps    6;

Qy      19 DTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKE 78
           |||     |:|:| | |   : ||| |:|:|||          ::  ||||:    |   :
Db     144 DTDV----PYLLLAVKSEPGRFAERQAVRETWGS----PAPGIRLLFLLGSPVGEA-GPD 194

Qy      79 VDQ----ESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVD 134
             :|     ||:|: |::  ||||| :| ||| :: : |: | ||  :||::   | |::
Db     195 LDSLVAWESRRYSDLLLWDFLDVPFNQTLKDLLLLAWLGRHCPTVSFVLRAQDDAFVHTP 254

Qy     135 YLTELL--LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFS 192
           |   |  |         : | :     |:|:|   ::| :| :    || : || ||| :
Db     255 ALLAHLRALPPASARSLYLGEVFTQAMPLRKPGGPFYVPESFFE-GGYPAYASGGGYVIA 313

Qy     193 GDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIV 252
           | :|  :   :  |     |||: |||:  | :  :      | :|       | || ::
Db     314 GRLAPWLLRAAARVAPFPFEDVYTGLCIRALGLVPQAHPGFLTAWPADRTADHCAFRNLL 373

Qy     253 ACHFIKPRTLLDYWQALENSR 273
               : |:  :  |: |:: |
Db     374 LVRPLGPQASIRLWKQLQDPR 394


RESULT 6
US-09-459-133-13
; Sequence 13, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:. Jaspers, Stephen R.
;  APPLICANT:  Gao, Zeren
;  TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
;  FILE REFERENCE: 98-77
;  CURRENT APPLICATION NUMBER: US/09/459,133
;  CURRENT FILING DATE:  1999-12-10
;  PRIOR APPLICATION NUMBER: 60/111,697
```

```
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
;   LENGTH: 389
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-459-133-13

  Query Match             20.4%;  Score 305;  DB 4;  Length 389;
  Best Local Similarity   31.0%;  Pred. No. 2.6e-28;
  Matches   77;  Conservative   41;  Mismatches  120;  Indels   10;  Gaps    5;

Qy      27 PFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE---TKEVDQES 83
           |:|:| | |     | | |:|:|||    | | :|   ||||:            |  ||
Db     140 PYLLLAVKSEPGHFAARQAVRETWGSP--VAGTRL--LFLLGSPLGMGGPDLRSLVTWES 195

Qy      84 QRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELL--L 141
           :|:||::  ||||| || ||| :: : |:   ||   ||::   | |:::  | | |  |
Db     196 RRYGDLLLWDFLDVPYNRTLKDLLLLTWLSHHCPDVNFVLQVQDDAFVHIPALLEHLQTL 255

Qy     142 KKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN 201
                 : | :     |:|:|   ::| |: :  | || : || ||| || :|  :
Db     256 PPTWARSLYLGEIFTQAKPLRKPGGPFYVPKTFFEGD-YPAYASGGGYVISGRLAPWLLQ 314

Qy     202 VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRT 261
           :  |   .  :||: | |   | :          | :|       |  | ::  | : |:
Db     315 AAARVAPFPFDDVYTGFCFRALGLAPRAHPGFLTAWPAERTRDPCAVRGLLLVHPVSPQD 374

Qy     262 LLDYWQAL 269
           :  |: |
Db     375 TIWLWRHL 382


RESULT 7
US-08-207-904-10
; Sequence 10, Application US/08207904
; Patent No. 5477002
;  GENERAL INFORMATION:
;    APPLICANT:  Tuttle, AnnMarie
;    APPLICANT:  Crossland, Lyle D.
;    TITLE OF INVENTION:  Anther-Specific cDNA Sequences, Genomic
;    TITLE OF INVENTION:  DNA Sequences and Recombinant DNA Sequences
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/207,904
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/908,242
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lazar, Steven R.
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  CGC 1624
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8615
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  372 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-207-904-10

  Query Match             9.2%;  Score 138;  DB 1;  Length 372;
  Best Local Similarity   23.7%;  Pred. No. 4.7e-08;
  Matches   53;  Conservative   41;  Mismatches   86;  Indels   44;  Gaps    9;

Qy      28 FLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTF---------FLLG--TTSSAAET 76
           |:|: : ::      | ::|:||    | || :|:           |::|   |
Db     103 FIVVGINTAFSSRKRRDSLRETW----MPKGDKLRKLEKEKGIVIRFVIGHSATRGGVLD 158

Qy      77 KEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIE-----WVHRFCPQAAFVMKTDSDMFI 131
           : :| |  :: | :: | :: |: |: || :        |       | | :| | |: :
Db     159 RAIDSEDAQYKDFLRLDHVEGYHELSTKTRLYFSKAVSIW------DADFYVKVDDDVHL 212

Qy     132 NVDYLTELLLKKNRTTRFFTGFLK----LNEFPIRQPFSKWFVSKSEYPW------DRYP 181
           |:  |   | |     | : | :|    |::  :|        : || |      ::|
Db     213 NLGMLANTLAKYKSKPRVYIGCMKSGPVLSQKGVR-------YYEPEY-WKFGEEGNKYF 264

Qy     182 PFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNI 225
           :|  |  | |:|| :   |  :     ||| :|  |  | :
Db     265 RHATGQIYGISRDLASYISINSGILHRYANEDVSLGSWLIGLEV 308


RESULT 8
PCT-US93-03076-2
; Sequence 2, Application PC/TUS9303076
;  GENERAL INFORMATION:
;    APPLICANT:  Whitehead Institute for Biomedical Research
;    TITLE OF INVENTION:  GAP-Associated Protein p190 and
;    TITLE OF INVENTION:  Transduction
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03076
;      FILING DATE:  19930331
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI92-03A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1513 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-03076-2

  Query Match             5.4%;  Score 81.5;  DB 5;  Length 1513;
  Best Local Similarity   23.6%;  Pred. No. 3;
  Matches   37;  Conservative   22;  Mismatches   63;  Indels   35;  Gaps    5;

Qy      42 ERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNL 101
           ::: : | : :::|  || |   ||||   |    :  |          |  |:   ||
Db     137 DQLGLEQDFEQKQMPDGKLLVDGFLLGIDVSRGMNRNFDD---------QLKFVSNLYNQ 187
```

```
Qy    102 TLKTMMGI--------EWVHRFCPQA---------AFVMKTDSDMFINVDY----LTELL 140
Db    188 LAKTKKPIVIVLTKCDEGVERYIRDAHTFALSKKNLQVVETSARSNVNVDLAFTTLVQLI 247

Qy    141 LKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db    248 DKSRGKTKIIPYFEAL-----KQQSQQIATAKDKYEW 279


RESULT 9
US-09-212-971-14
; Sequence 14, Application US/09212971B
; Patent No. 6107041
; GENERAL INFORMATION:
;  APPLICANT: Korneluk, Robert G
;  APPLICANT: MacKenzie, Alexander E
;  APPLICANT: Liston, Peter
;  APPLICANT: Baird, Stephen
;  APPLICANT: Tsang, Benjamin K
;  APPLICANT: Pratt, Christine
;  TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
;  TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
;  TITLE OF INVENTION: DISEASE
;  FILE REFERENCE: 07891/009002
;  CURRENT APPLICATION NUMBER: US/09/212,971B
;  CURRENT FILING DATE: 1998-12-16
;  EARLIER APPLICATION NUMBER: 60/017,354
;  EARLIER FILING DATE: 1996-04-26
;  EARLIER APPLICATION NUMBER: 60/030,590
;  EARLIER FILING DATE: 1996-11-14
;  EARLIER APPLICATION NUMBER: 08/800,929
;  EARLIER FILING DATE: 1997-02-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 612
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-212-971-14

  Query Match             5.3%;  Score 79.5;  DB 3;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.3;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy     29 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 78
Db    382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy     79 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 131
Db    442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy    132 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 183
Db    492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK---------------------------- 523

Qy    184 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
Db    524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy    241 LRFSVCLFRRI 251
Db    565 --CKVCMDREV 573


RESULT 10
US-08-800-929A-14
; Sequence 14, Application US/08800929A
; Patent No. 6133437
;  GENERAL INFORMATION:
;    APPLICANT:  Korneluk, Robert G
;    APPLICANT:  MacKenzie, Alexander E
;    APPLICANT:  Liston, Peter
;    APPLICANT:  Baird, Stephen
;    APPLICANT:  Tsang, Benjamin K
;    APPLICANT:  Pratt, Christine
;    TITLE OF INVENTION:  DETECTION AND MODULATION OF
;    TITLE OF INVENTION:  IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERA
;    TITLE OF INVENTION:  DISEASE
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Clark & Elbing LLP
;      STREET:  176 Federal Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/800,929A
;      FILING DATE:  13-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/030,590
;      FILING DATE:  14-NOV-1996
;      APPLICATION NUMBER:  60/017,354
;      FILING DATE:  26-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bieker-Brady, Kristina
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:  07891/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-428-0200
;      TELEFAX:  617-428-7045
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-800-929A-14

  Query Match             5.3%;  Score 79.5;  DB 4;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.3;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy     29 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 78
Db    382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy     79 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 131
Db    442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy    132 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 183
Db    492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK---------------------------- 523

Qy    184 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
Db    524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy    241 LRFSVCLFRRI 251
Db    565 --CKVCMDREV 573


RESULT 11
US-08-569-749-14
; Sequence 14, Application US/08569749
```

```
; Patent No. 6187557
;  GENERAL INFORMATION:
;    APPLICANT:  Rothe, Mike
;    APPLICANT:  Goeddel, David V
;    TITLE OF INVENTION:  INHIBITORS OF APOPTOSIS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/569,749
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brezner, David J.
;      REGISTRATION NUMBER:  24,774
;      REFERENCE/DOCKET NUMBER:  A-62464/DJB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)781-1989
;      TELEFAX:  (415)398-3249
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-569-749-14

  Query Match             5.3%;  Score 79.5;  DB 4;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.3;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      29 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 78
           :|:: |   |   |    |  :|||  :: :  |:  :|        |         | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy      79 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 131
             |    ||  :|:|: : ::  ||     |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     132 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 183
                  : :  :|:| |     |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK---------------------------- 523

Qy     184 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
                  :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     241 LRFSVCLFRRI 251
             ||: | :
Db     565 --CKVCMDREV 573


RESULT 12
US-09-617-053A-14
; Sequence 14, Application US/09617053A
; Patent No. 6300492
; GENERAL INFORMATION:
;  APPLICANT: Korneluk, Robert G
;  APPLICANT:  MacKenzie, Alexander E
;  APPLICANT:  Liston, Peter
;  APPLICANT:  Baird, Stephen
;  APPLICANT:  Tsang, Benjamin K
;  APPLICANT:  Pratt, Christine
;  TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
;  TITLE OF INVENTION:  NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
;  TITLE OF INVENTION:  DISEASE
;  FILE REFERENCE: 07891/009003
;  CURRENT APPLICATION NUMBER: US/09/617,053A
;  CURRENT FILING DATE:  2000-07-14
;  PRIOR APPLICATION NUMBER: US 08/800,929
;  PRIOR FILING DATE: 1997-02-13
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 612
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-617-053A-14

  Query Match             5.3%;  Score 79.5;  DB 4;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.3;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      29 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 78
           :|:: |   |   |    |  :|||  :: :  |:  :|        |         | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy      79 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 131
             |    ||  :|:|: : ::  ||     |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     132 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 183
                  : :  :|:| |     |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK---------------------------- 523

Qy     184 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
                  :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     241 LRFSVCLFRRI 251
             ||: | :
Db     565 --CKVCMDREV 573


RESULT 13
PCT-US96-12860-14
; Sequence 14, Application PC/TUS9612860
;  GENERAL INFORMATION:
;    APPLICANT:  TULARIK, INC.
;    TITLE OF INVENTION:  INHIBITORS OF APOPTOSIS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/12860
;      FILING DATE:  06 AUG 1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. Serial Nos. 08/512,946 & 08/569,749
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brezner, David J.
;      REGISTRATION NUMBER:  24,774
;      REFERENCE/DOCKET NUMBER:  A-62464/DJB
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (415)781-1989
;      TELEFAX:  (415)398-3249
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US96-12860-14

  Query Match             5.3%;  Score 79.5;  DB 5;  Length 612;
  Best Local Similarity   19.9%;  Pred. No. 1.3;
  Matches   50;  Conservative   35;  Mismatches   79;  Indels   87;  Gaps   10;

Qy      29 LVLLVTSSHKQLAE----RMAIRQTWGKERMVKGKQLKT------FFLLGTTSSAAETKE 78
           :|:: |   |   |    |  :|||  :: :  |:  :|        |         | ||
Db     382 VVMMSTPVVKAALEMGFSRSLVRQTVQRQILATGENYRTVNDIVSVLLNAEDERREEEKE 441

Qy      79 VDQESQRHGD--IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFI 131
             |    ||  :|:|: : ::  ||     |  ::          :|: : | : |:
Db     442 RQTEEMASGDLSLIRKNRMALFQQLTHVLPILDNLL----------EASVITKQEHDIIR 491

Qy     132 NV--------DYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPF 183
                    : :  :|:| |     |   ||
Db     492 QKTQIPLQARELIDTVLVKGNAAANIFKNSLK---------------------------- 523

Qy     184 CSGTGYVFSGDVASQVYN---VSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG 240
                  :: | :|    | |:: ||  |||  || ||    ||:|  :
Db     524 ----------EIDSTLYENLFVEKNMKYIPTEDV-SGLSLEEQLRRLQEERT-------- 564

Qy     241 LRFSVCLFRRI 251
               ||: | :
Db     565 --CKVCMDREV 573


RESULT 14
PCT-US95-05922A-2
; Sequence 2, Application PC/TUS9505922A
;  GENERAL INFORMATION:
;    APPLICANT:  HE, ET AL.
;    TITLE OF INVENTION:  Human Inhibitor of Apoptosis Gene 1
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/05922A
;      FILING DATE:  11 MAY 1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-292
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  438 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
PCT-US95-05922A-2

  Query Match             5.2%;  Score 78;  DB 5;  Length 438;
  Best Local Similarity   19.3%;  Pred. No. 1.2;
  Matches   52;  Conservative   46;  Mismatches   88;  Indels   84;  Gaps   11;

Qy       2 KEQSFVYKKDGNF-------LKLPDTDCRQT--PPFL-----------VLLVTSSHKQL 40
           | | || :  | :        |   ||   :   || :           |:: |   |
Db     160 KGQEFVDEIQGRYPHLLEQLLSTSDTTGEENADPPIIHFGPGESSSEDAVMMNTPVVKSA 219

Qy      41 AE----RMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGD-------- 88
             |    |  ::||   : :  |:  ||   : :    || :: ::| ::  :
Db     220 LEMGFNRDLVKQTVQSKILTTGENYKTVNDIVSALLNAEDEKREEEKEKQAEEMASDDLS 279

Qy      89 IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFINV-DYLTELLLK 142
           :|:|: : ::  ||     |  ::    :::   |   ::|  : : :   : :  :|:|
Db     280 LIRKNRMALFQQLTCVLPILDNLLKANVINK---QEHDIIKQKTQIPLQARELIDTILVK 336

Qy     143 KNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN- 201
           |     |   ||                                          :: | :|
Db     337 GNAAANIFKNCLK-------------------------------------EIDSTLYKN 358

Qy     202 --VSKSVPYIKLEDVFVGLCLERLNIRLEE 229
             | |:: ||  |||  || ||    ||:|
Db     359 LFVDKNMKYIPTEDV-SGLSLEEQLRRLQE 387


RESULT 15
US-08-569-749-2
; Sequence 2, Application US/08569749
; Patent No. 6187557
;  GENERAL INFORMATION:
;    APPLICANT:  Rothe, Mike
;    APPLICANT:  Goeddel, David V
;    TITLE OF INVENTION:  INHIBITORS OF APOPTOSIS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/569,749
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brezner, David J.
;      REGISTRATION NUMBER:  24,774
;      REFERENCE/DOCKET NUMBER:  A-62464/DJB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)781-1989
;      TELEFAX:  (415)398-3249
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  618 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-569-749-2

  Query Match             5.2%;  Score 78;  DB 4;  Length 618;
```

```
  Best Local Similarity   19.3%;  Pred. No. 2;
  Matches   52;  Conservative   46;  Mismatches   88;  Indels   84;  Gaps   11;

Qy       2 KEQSFVYKKDGNF-------LKLPDTDCRQT--PPFL------------VLLVTSSHKQL 40
           | | || :  | :       |   ||   :   || :            |:: |   |
Db     340 KGQEFVDEIQGRYPHLLEQLLSTSDTTGEENADPPIIHFGPGESSSEDAVMMNTPVVKSA 399

Qy      41 AE----RMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGD-------- 88
            |    |  ::||   : :  |:  ||   : :    || :: ::| ::  :
Db     400 LEMGFNRDLVKQTVQSKILTTGENYKTVNDIVSALLNAEDEKREEEKEKQAEEMASDDLS 459

Qy      89 IIQKDFLDVYYNLT-----LKTMMGIEWVHRFCPQAAFVMKTDSDMFINV-DYLTELLLK 142
           :|:|: : ::  ||     |  ::    :::   |   ::|  : : :   : :  :|:|
Db     460 LIRKNRMALFQQLTCVLPILDNLLKANVINK---QEHDIIKQKTQIPLQARELIDTILVK 516

Qy     143 KNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYN- 201
             |     |   ||                                    :: | :|
Db     517 GNAAANIFKNCLK-----------------------------------EIDSTLYKN 538

Qy     202 --VSKSVPYIKLEDVFVGLCLERLNIRLEE 229
             | |:: ||  |||  || ||    ||:|
Db     539 LFVDKNMKYIPTEDV-SGLSLEEQLRRLQE 567


Search completed: April  4, 2003, 14:45:52
Job time : 18 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 14:45:11 ; Search time 37 Seconds
(without alignments)
462.651 Million cell updates/sec

Title: US-09-914-152-1_COPY_31_310
Perfect score: 1498
Sequence: 1 FKEQSFVYKKDGNFLKLPDT.....TLLDYWQALENSRGEDCPPV 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	577	38.5	326	10	US-09-804-357-8	Sequence 8, Appli
2	577	38.5	326	10	US-09-804-006-8	Sequence 8, Appli
3	569	38.0	422	10	US-09-804-357-11	Sequence 11, Appl
4	569	38.0	422	10	US-09-529-063-25	Sequence 25, Appl
5	569	38.0	422	10	US-09-804-006-11	Sequence 11, Appl
6	567	37.9	422	9	US-10-112-616A-2	Sequence 2, Appli
7	555.5	37.1	331	10	US-09-804-357-9	Sequence 9, Appli
8	555.5	37.1	331	10	US-09-804-006-9	Sequence 9, Appli
9	548.5	36.6	319	10	US-09-739-451-2	Sequence 2, Appli
10	548.5	36.6	320	10	US-09-739-451-10	Sequence 10, Appl
11	548.5	36.6	331	9	US-09-992-598-209	Sequence 209, App
12	548.5	36.6	331	9	US-09-989-293A-209	Sequence 209, App
13	548.5	36.6	331	9	US-10-063-547-36	Sequence 36, Appl
14	548.5	36.6	331	9	US-09-989-735-209	Sequence 209, App
15	548.5	36.6	331	9	US-09-990-444-209	Sequence 209, App
16	548.5	36.6	331	9	US-09-989-730-209	Sequence 209, App
17	548.5	36.6	331	9	US-09-990-436-209	Sequence 209, App
18	548.5	36.6	331	9	US-09-991-181-209	Sequence 209, App
19	548.5	36.6	331	9	US-09-993-687-209	Sequence 209, App
20	548.5	36.6	331	9	US-09-989-734-209	Sequence 209, App
21	548.5	36.6	331	9	US-09-997-653-209	Sequence 209, App
22	548.5	36.6	331	9	US-10-174-590-162	Sequence 162, App
23	548.5	36.6	331	9	US-10-176-758-162	Sequence 162, App
24	548.5	36.6	331	9	US-10-063-616-36	Sequence 36, Appl
25	548.5	36.6	331	9	US-10-175-737-162	Sequence 162, App
26	548.5	36.6	331	9	US-09-993-667-209	Sequence 209, App
27	548.5	36.6	331	9	US-10-063-502-36	Sequence 36, Appl
28	548.5	36.6	331	9	US-10-173-706-162	Sequence 162, App
29	548.5	36.6	331	9	US-10-175-738-162	Sequence 162, App
30	548.5	36.6	331	9	US-10-175-752-162	Sequence 162, App
31	548.5	36.6	331	9	US-10-176-482-162	Sequence 162, App
32	548.5	36.6	331	9	US-10-176-757-162	Sequence 162, App
33	548.5	36.6	331	9	US-10-176-913-162	Sequence 162, App
34	548.5	36.6	331	9	US-10-180-552-162	Sequence 162, App
35	548.5	36.6	331	9	US-10-180-557-162	Sequence 162, App
36	548.5	36.6	331	9	US-09-990-438-209	Sequence 209, App
37	548.5	36.6	331	9	US-09-990-562-209	Sequence 209, App
38	548.5	36.6	331	9	US-09-997-428-209	Sequence 209, App
39	548.5	36.6	331	9	US-09-997-666-209	Sequence 209, App
40	548.5	36.6	331	9	US-10-173-700-162	Sequence 162, App
41	548.5	36.6	331	9	US-10-174-572-162	Sequence 162, App
42	548.5	36.6	331	9	US-10-174-579-162	Sequence 162, App
43	548.5	36.6	331	9	US-10-174-582-162	Sequence 162, App
44	548.5	36.6	331	9	US-10-174-588-162	Sequence 162, App
45	548.5	36.6	331	9	US-10-175-739-162	Sequence 162, App

ALIGNMENTS

RESULT 1
US-09-804-357-8
; Sequence 8, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Zhou, Jianghong
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: LEPTIN INDUCED GENES
; FILE REFERENCE: 07334/109001
; CURRENT APPLICATION NUMBER: US/09/804,357
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/195,896
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: US 60/108,379
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: US 09/150,857
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-804-357-8

Query Match 38.5%; Score 577; DB 10; Length 326;
Best Local Similarity 42.8%; Pred. No. 7.1e-50;
Matches 116; Conservative 46; Mismatches 93; Indels 16; Gaps 4;

```
Qy      14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
           ||      | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db      65 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 124

Qy      74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
              : |:||||   ||| :||:| |:||||||:||: ||  || :| :||||||:|:|:
Db     125 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 184

Qy     134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
           | |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
Db     185 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 243
```

```
Qy    192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
          | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db    244 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 295

Qy    247 LFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
           :||::  | | |  :   |  : : :   |
Db    296 RYRRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 2
US-09-804-006-8
; Sequence 8, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
;  APPLICANT: White, David
;  APPLICANT:  Zhou, Jianghong
;  APPLICANT:  Tartaglia, Louis A.
;  TITLE OF INVENTION: LEPTIN INDUCED GENES
;  FILE REFERENCE: 07334/126001
;  CURRENT APPLICATION NUMBER: US/09/804,006
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/292,228
;  PRIOR FILING DATE: 1999-04-15
;  PRIOR APPLICATION NUMBER: US 60/108,379
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 09/150,857
;  PRIOR FILING DATE: 1998-09-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 326
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-804-006-8

  Query Match             38.5%;  Score 577;  DB 10;  Length 326;
  Best Local Similarity   42.8%;  Pred. No. 7.1e-50;
  Matches  116;  Conservative   46;  Mismatches   93;  Indels   16;  Gaps    4;

Qy     14 FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSA 73
          ||      | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db     65 FLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADP 124

Qy     74 AETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINV 133
             : |:||||   ||| :||:| |:||||||:||: ||  || :| :|||||||:|:|:
Db    125 VLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNM 184

Qy    134 DYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVF 191
          | |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:|
Db    185 DNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIF 243

Qy    192 SGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVC 246
          | |||  :|  |     : ||||:||||| :| |          |   |     : :|:|
Db    244 SADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLC 295

Qy    247 LFRRIVACHFIKPRTLLDYWQALENSRGEDC 277
           :||::  | | |  :   |  : : :   |
Db    296 RYRRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 3
US-09-804-357-11
; Sequence 11, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
;  APPLICANT: White, David
;  APPLICANT:  Zhou, Jianghong
;  APPLICANT:  Tartaglia, Louis A.
;  TITLE OF INVENTION: LEPTIN INDUCED GENES
;  FILE REFERENCE: 07334/109001
;  CURRENT APPLICATION NUMBER: US/09/804,357
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/195,896
;  PRIOR FILING DATE: 1998-11-19
;  PRIOR APPLICATION NUMBER: US 60/108,379
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 09/150,857
;  PRIOR FILING DATE: 1998-09-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-804-357-11

  Query Match             38.0%;  Score 569;  DB 10;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 6.3e-49;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy     22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
          |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db    146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
           :||::: ||||:::|| ||||||:||:||: ||  :||   :||||||||:| :||
Db    206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy    140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
          ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||:|
Db    266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db    326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    386 QPSELIKYWNHLQQNKHNAC 405


RESULT 4
US-09-529-063-25
; Sequence 25, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
;  APPLICANT: FUKUSHIMA, DAIKICHI
;  APPLICANT:  SHIBAYAMA, SHIRO
;  APPLICANT:  TADA, HIDEAKI
;  TITLE OF INVENTION: POLYPEPTIDE, cDNA ENCODING THE POLYPEPTIDE, AND USE OF
;  TITLE OF INVENTION:  THE BOTH
;  FILE REFERENCE: Q58769
;  CURRENT APPLICATION NUMBER: US/09/529,063
;  CURRENT FILING DATE:  2000-04-07
;  PRIOR APPLICATION NUMBER: PCT/JP98/04514
;  PRIOR FILING DATE: 1998-10-06
;  PRIOR APPLICATION NUMBER: JP 9-274674
;  PRIOR FILING DATE: 1997-10-07
;  NUMBER OF SEQ ID NOS: 117
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-529-063-25

  Query Match             38.0%;  Score 569;  DB 10;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 6.3e-49;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy     22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
          |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db    146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
           :||::: ||||:::|| ||||||:||:||: ||  :||   :||||||||:| :||
Db    206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265
```

```
Db    206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy    140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
          ||| :   |  :|||:|     | |   |||::    || :||| |||||||||||:|
Db    266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| |      ::  |   : :| | :  ::  |
Db    326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    386 QPSELIKYWNHLQQNKHNAC 405


RESULT 5
US-09-804-006-11
; Sequence 11, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
;  APPLICANT: White, David
;  APPLICANT:  Zhou, Jianghong
;  APPLICANT:  Tartaglia, Louis A.
;  TITLE OF INVENTION: LEPTIN INDUCED GENES
;  FILE REFERENCE: 07334/126001
;  CURRENT APPLICATION NUMBER: US/09/804,006
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/292,228
;  PRIOR FILING DATE: 1999-04-15
;  PRIOR APPLICATION NUMBER: US 60/108,379
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 09/150,857
;  PRIOR FILING DATE: 1998-09-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-804-006-11

  Query Match             38.0%;  Score 569;  DB 10;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 6.3e-49;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy     22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
          |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db    146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
           :||::: ||||:::|| |||||:||:||: ||  :||   :||||||||:| :||
Db    206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy    140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
          ||| :   |  :|||:|     | |   |||::    || :||| |||||||||||:|
Db    266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| |      ::  |   : :| | :  ::  |
Db    326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    386 QPSELIKYWNHLQQNKHNAC 405


RESULT 6
US-10-112-616A-2
; Sequence 2, Application US/10112616A
; Publication No. US20030005477A1
; GENERAL INFORMATION:
;  APPLICANT: Leviten, Michael W.
;  TITLE OF INVENTION: TRANSGENIC MICE CONTAINING BETA 3 GalT2
;  TITLE OF INVENTION:  GENE DISRUPTIONS
;  FILE REFERENCE: R-031
;  CURRENT APPLICATION NUMBER: US/10/112,616A
;  CURRENT FILING DATE:  2002-09-05
;  PRIOR APPLICATION NUMBER: US 60/280,362
;  PRIOR FILING DATE: 2001-03-29
;  PRIOR APPLICATION NUMBER: US 60/326,700
;  PRIOR FILING DATE: 2001-10-02
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-112-616A-2

  Query Match             37.9%;  Score 567;  DB 9;  Length 422;
  Best Local Similarity   41.2%;  Pred. No. 1e-48;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy     22 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 79
          |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :    : :
Db    146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 205

Qy     80 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 139
           :||::: ||||:::|| |||||:||:||: ||  :||   :||||||||:| :||
Db    206 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 265

Qy    140 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 197
          ||| :   |  :|||:|     | |   |||::    || :||| |||||||||||:|
Db    266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy    198 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 257
          ::: ||  :  : ||||:||:|| :| |      ::  |   : :| | :  ::  |
Db    326 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy    258 KPRTLLDYWQALENSRGEDC 277
          :|  |: ||  |: ::   |
Db    386 QPSELIKYWNHLQQNKHNAC 405


RESULT 7
US-09-804-357-9
; Sequence 9, Application US/09804357
; Patent No. US20010024808A1
; GENERAL INFORMATION:
;  APPLICANT: White, David
;  APPLICANT:  Zhou, Jianghong
;  APPLICANT:  Tartaglia, Louis A.
;  TITLE OF INVENTION: LEPTIN INDUCED GENES
;  FILE REFERENCE: 07334/109001
;  CURRENT APPLICATION NUMBER: US/09/804,357
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/195,896
;  PRIOR FILING DATE: 1998-11-19
;  PRIOR APPLICATION NUMBER: US 60/108,379
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 09/150,857
;  PRIOR FILING DATE: 1998-09-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-804-357-9

  Query Match             37.1%;  Score 555.5;  DB 10;  Length 331;
  Best Local Similarity   39.3%;  Pred. No. 1e-47;
  Matches  107;  Conservative   51;  Mismatches  111;  Indels    3;  Gaps    1;

Qy      1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
            | |   :|::|  |     ::|    ||||:||||    :  | ||| |||:::   | :
```

```
Db      52 FYEYEPIYRQDFRFTLREHSNCSHQNPFLVILVTSRPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
           : |||||  :   :      :: |   :||||::|||| | ||||||:|   ||  |||
Db     112 VLTFFLLGQQAEREDKTLALSLEDEHVLYGDIIRQDFLDTYNNLTLKTIMAFRWVMEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
            | ::||||:|:|||   | : ||  | : :||||:  :: :  |  | |  :|  |||:
Db     172 NAKYIMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFFHKNHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
             :||:||| ||: |||:  :|| :   |  || |||:||:||  | : :         ||
Db     232 KVFPPYCSGLGYIMSGDLVPRVYEMMSHVKPIKFEDVYVGICLNLLKVDIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
              :   ||  ||::| |    : :: :|| :
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 8
US-09-804-006-9
; Sequence 9, Application US/09804006
; Patent No. US20020119517A1
; GENERAL INFORMATION:
;  APPLICANT: White, David
;  APPLICANT:  Zhou, Jianghong
;  APPLICANT:  Tartaglia, Louis A.
;  TITLE OF INVENTION: LEPTIN INDUCED GENES
;  FILE REFERENCE: 07334/126001
;  CURRENT APPLICATION NUMBER: US/09/804,006
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/292,228
;  PRIOR FILING DATE: 1999-04-15
;  PRIOR APPLICATION NUMBER: US 60/108,379
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: US 09/150,857
;  PRIOR FILING DATE: 1998-09-10
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-804-006-9

  Query Match             37.1%;  Score 555.5;  DB 10;  Length 331;
  Best Local Similarity   39.3%;  Pred. No. 1e-47;
  Matches  107;  Conservative   51;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           | |   :|::|  |     ::|    ||||:||||     :  | ||| |||:::   | :
Db      52 FYEYEPIYRQDFRFTLREHSNCSHQNPFLVILVTSRPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
           : ||||||  :   :      :: |   :||||::|||| | ||||||:|   ||  |||
Db     112 VLTFFLLGQQAEREDKTLALSLEDEHVLYGDIIRQDFLDTYNNLTLKTIMAFRWVMEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
            | ::||||:|:|||   | : ||  | : :||||:  :: :  |  | |  :|  |||:
Db     172 NAKYIMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFFHKNHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
             :||:||| ||: |||:  :|| :   |  || |||:||:||  | : :         ||
Db     232 KVFPPYCSGLGYIMSGDLVPRVYEMMSHVKPIKFEDVYVGICLNLLKVDIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
              :   ||  ||::| |    : :: :|| :
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 9
US-09-739-451-2
; Sequence 2, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
;  APPLICANT: Ebner, Reinhard
;  TITLE OF INVENTION: Dendriac and Brainiac-3
;  FILE REFERENCE: PF464
;  CURRENT APPLICATION NUMBER: US/09/739,451
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: 09/213,364
;  PRIOR FILING DATE: 1998-12-17
;  PRIOR APPLICATION NUMBER: 60/077,687
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/108,928
;  PRIOR FILING DATE: 1998-11-17
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 319
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-739-451-2

  Query Match             36.6%;  Score 548.5;  DB 10;  Length 319;
  Best Local Similarity   38.6%;  Pred. No. 4.9e-47;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           | |   :|::| :|     ::|    ||||:||||     :  | ||| |||:::   | :
Db      40 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 99

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
           : ||||||  :   :      :: |   :||||::|||| | ||||||:|   ||  |||
Db     100 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 159

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
            | :|||||:|:|||   | : ||  | : :||||:  :: :  |  : |  :|  |||:
Db     160 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 219

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
             :||:||| ||: | |:  ::| :   |  || |||:||:||  | : :         ||
Db     220 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 279

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
              :   ||  ||::| |    : :: :|| :
Db     280 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 311


RESULT 10
US-09-739-451-10
; Sequence 10, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
;  APPLICANT: Ebner, Reinhard
;  TITLE OF INVENTION: Dendriac and Brainiac-3
;  FILE REFERENCE: PF464
;  CURRENT APPLICATION NUMBER: US/09/739,451
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER: 09/213,364
;  PRIOR FILING DATE: 1998-12-17
;  PRIOR APPLICATION NUMBER: 60/077,687
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/108,928
;  PRIOR FILING DATE: 1998-11-17
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 320
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-739-451-10

  Query Match             36.6%;  Score 548.5;  DB 10;  Length 320;
  Best Local Similarity   38.6%;  Pred. No. 5e-47;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;
```

```
Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      40 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 99

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     100 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 159

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     160 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 219

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     220 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 279

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     280 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 311


RESULT 11
US-09-992-598-209
; Sequence 209, Application US/09992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C20
;  CURRENT APPLICATION NUMBER: US/09/992,598
;  CURRENT FILING DATE:  2001-11-14
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088810
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088858
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088861
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088876
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089440
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089512
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089532
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089538
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089598
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089599
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089600
;  PRIOR FILING DATE: 1998-06-17
```

; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 36.6%; Score 548.5; DB 9; Length 331;
Best Local Similarity 38.6%; Pred. No. 5.2e-47;
Matches 105; Conservative 53; Mismatches 111; Indels 3; Gaps 1;

```
Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323
```

RESULT 12
US-09-989-293A-209
; Sequence 209, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi,Avi J.
; APPLICANT: Baker,Kevin P.
; APPLICANT: Botstein,David
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Fong,Sherman
; APPLICANT: Gerber,Hanspeter
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,J.Christopher
; APPLICANT: Gurney,Austin L.
; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C66
; CURRENT APPLICATION NUMBER: US/09/989,293A
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13

; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02

```
;  PRIOR APPLICATION NUMBER: 60/091544
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091519
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091626
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091633
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091978
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/091982
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/092182
;  PRIOR FILING DATE: 1998-07-09

  Query Match             36.6%;  Score 548.5;  DB 9;  Length 331;
  Best Local Similarity   38.6%;  Pred. No. 5.2e-47;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           | |   :|::| :|     ::|    ||||:||||    :  | ||| |||:::   | :
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
             : ||||||  :   :      :: |   :||||::|||| | ||||||:|   ||  |||
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
             | :|||||:|:|||   | : ||  | : :||||:  :: :  |  : |  :|  |||:
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
             :||:||| ||: | |:  ::| :   |  || |||:||:||  | : :         ||
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
             :   ||  ||::| |    : :: :|| :
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 13
US-10-063-547-36
; Sequence 36, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,547
;  CURRENT FILING DATE:  2002-05-02
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 36
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-547-36

  Query Match             36.6%;  Score 548.5;  DB 9;  Length 331;
  Best Local Similarity   38.6%;  Pred. No. 5.2e-47;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           | |   :|::| :|     ::|    ||||:||||    :  | ||| |||:::   | :
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
             : ||||||  :   :      :: |   :||||::|||| | ||||||:|   ||  |||
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
             | :|||||:|:|||   | : ||  | : :||||:  :: :  |  : |  :|  |||:
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
             :||:||| ||: | |:  ::| :   |  || |||:||:||  | : :         ||
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
             :   ||  ||::| |    : :: :|| :
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 14
US-09-989-735-209
; Sequence 209, Application US/09989735
; Publication No. US20020193299A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C61
;  CURRENT APPLICATION NUMBER: US/09/989,735
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
```

; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

```
;  PRIOR APPLICATION NUMBER: 60/092182
;  PRIOR FILING DATE: 1998-07-09

  Query Match             36.6%;  Score 548.5;  DB 9;  Length 331;
  Best Local Similarity   38.6%;  Pred. No. 5.2e-47;
  Matches  105;  Conservative   53;  Mismatches  111;  Indels    3;  Gaps    1;

Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
           | |   :|:::| :|     ::|    ||||:|||    :  | ||| |||:::   | :
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
           : ||||||  :   :       :: |   :||||::|||| | ||||||:|   ||  |||
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
            | :|||||:|:|||   | : ||  | : :||||:  :: :  |  : |  :|  |||:
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
             :||:||| ||: | |:  ::| :   |  || |||:||:||  | : :         ||
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
             :   ||  ||::| |    : :: :|| :
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323


RESULT 15
US-09-990-444-209
; Sequence 209, Application US/09990444
; Publication No. US20020193300A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C19
;  CURRENT APPLICATION NUMBER: US/09/990,444
;  CURRENT FILING DATE:  2001-11-14
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088810
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088858
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088861
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088876
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089440
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089512
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089532
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089538
;  PRIOR FILING DATE: 1998-06-17
```

; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 36.6%; Score 548.5; DB 9; Length 331;
Best Local Similarity 38.6%; Pred. No. 5.2e-47;
Matches 105; Conservative 53; Mismatches 111; Indels 3; Gaps 1;

```
Qy       1 FKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQ 60
Db      52 FYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYE 111

Qy      61 LKTFFLLGTTSSAAE---TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCP 117
Db     112 VLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCP 171

Qy     118 QAAFVMKTDSDMFINVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPW 177
Db     172 NAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPF 231

Qy     178 DRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFF 237
Db     232 KVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFF 291

Qy     238 PGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 269
Db     292 LYRIHLDVCQLRRVIAAHGFSSKEIITFWQVM 323
```

Search completed: April 4, 2003, 14:53:31
Job time : 38 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:22:18 ; Search time 5049.19 Seconds
(without alignments)
15994.713 Million cell updates/sec

Title: US-09-914-152-2
Perfect score: 2775
Sequence: 1 gtgaattcctctttctctgc.....aaaaaaaaaaaaaaaaaaa 2775

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		2775	100.0	2775	6	E38419	E38419 Novel polyp
2		2730	98.4	2762	9	AB020337	AB020337 Homo sapi
3		2332	84.0	170121	9	AF064860	AF064860 Homo sapi
4		2332	84.0	340000	9	HS21C080	AL163280 Homo sapi
5		2329.4	83.9	10562	6	E38420	E38420 Novel polyp
6		933	33.6	933	9	AF145784	AF145784 Homo sapi
7		924	33.3	933	9	HSA6078	AJ006078 Homo sapi
8		899.2	32.4	1576	9	AB041416	AB041416 Homo sapi
9		882.6	31.8	1570	9	AB041415	AB041415 Pan panis
10		879.6	31.7	1565	9	AB041413	AB041413 Homo sapi
11		872.4	31.4	1360	9	AB041412	AB041412 Gorilla g
12		869.6	31.3	1566	9	AB041414	AB041414 Pan trogl
13		850.4	30.6	1579	9	AB041417	AB041417 Pongo pyg
14		542.2	19.5	196900	2	AC020851	AC020851 Mus muscu
c 15		534.4	19.3	175861	2	AC120346	AC120346 Mus muscu
16		520.8	18.8	927	10	AF254738	AF254738 Mus muscu
c 17		341.8	12.3	65400	2	AC109264	AC109264 Mus muscu
18		185.2	6.7	917	10	AB039137	AB039137 Mus muscu
19		185.2	6.7	917	10	AB039140	AB039140 Mus muscu
20		185.2	6.7	917	10	AB039141	AB039141 Mus muscu
21		185.2	6.7	917	10	AB039142	AB039142 Mus muscu
22		183.6	6.6	917	10	AB039136	AB039136 Mus muscu
23		183.6	6.6	917	10	AB039138	AB039138 Mus muscu
24		183.6	6.6	917	10	AB039143	AB039143 Mus spici
25		182	6.6	917	10	AB039134	AB039134 Mus muscu
26		182	6.6	917	10	AB039135	AB039135 Mus muscu
27		182	6.6	917	10	AB039139	AB039139 Mus muscu
28		182	6.6	1175	10	AF029790	AF029790 Mus muscu
c 29		182	6.6	219368	10	AC098721	AC098721 Mus muscu
30		166	6.0	1053	9	AB041409	AB041409 Pan panis
31		165	5.9	1056	9	AB041410	AB041410 Gorilla g
32		165	5.9	1059	9	AB041411	AB041411 Pongo pyg
33		164.4	5.9	1060	9	AB041408	AB041408 Pan trogl
34		163.4	5.9	1037	9	AB041407	AB041407 Homo sapi
35		163.4	5.9	1739	6	E07739	E07739 cDNA encodi
36		163.4	5.9	202001	9	AC016723	AC016723 Homo sapi
37		161.8	5.8	1909	9	AF117222	AF117222 Homo sapi
c 38		151.8	5.5	137703	9	AC092036	AC092036 Homo sapi
39		151.8	5.5	159095	2	AC016935	AC016935 Homo sapi
c 40		148.8	5.4	161824	9	AC096721	AC096721 Homo sapi
41		142.8	5.1	996	10	AB039156	AB039156 Mus muscu
42		142.8	5.1	996	10	AB039158	AB039158 Mus muscu
43		142.8	5.1	996	10	AB039162	AB039162 Mus muscu
44		142.8	5.1	996	10	AB039163	AB039163 Mus spici
c 45		142	5.1	164076	9	AC008006	AC008006 Homo sapi

ALIGNMENTS

RESULT 1
E38419
LOCUS E38419 2775 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel polypeptide.
ACCESSION E38419
VERSION E38419.1 GI:18626993
KEYWORDS JP 2000245464-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2775)
AUTHORS Narimatsu,H., Isshiki,S., Togayauchi,A. and Sasaki,K.
TITLE Novel polypeptide
JOURNAL Patent: JP 2000245464-A 1 12-SEP-2000;

```
            KYOWA HAKKO KOGYO CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 2000245464-A/1
            PD   12-SEP-2000
            PF   25-FEB-1999 JP 1999047571
            PR
            PI   HISASHI NARIMATSU,SOICHIRO ISSHIKI,AKIRA TOGAYAUCHI, PI
            KATSUTOSHI SASAKI
            PC   C12N15/09,A01K67/027,C12N1/21,C12N5/10,C12N9/10,C12P19/00, PC
             C12P21/02,
            PC   C12P21/08,C12Q1/68,G01N33/53//(C12N1/21,C12R1:185),(C12N5/10,
            PC   C12R1:91),
            PC   (C12P21/02,C12R1:185),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
            PC   (C12N5/00,C12R1:91)
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (402). .(1331).

FEATURES              Location/Qualifiers
     source           1. .2775
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"

BASE COUNT       681 a    698 c    669 g    727 t
ORIGIN

  Query Match          100.0%;  Score 2775;  DB 6;  Length 2775;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2775;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCA 60

Qy      61 GAGCTGCAGGCTCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAGCTGCAGGCTCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTC 120

Qy     121 TCAGGCCTTTGGCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCAGGCCTTTGGCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTT 180

Qy     181 GGTCTTGGACTGAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGTCTTGGACTGAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCAC 240

Qy     241 GTGATTCCTGTCAGAATCACCATTTTTGGTAAACAAACCAAGCCCAGAACCTGATAATTA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGATTCCTGTCAGAATCACCATTTTTGGTAAACAAACCAAGCCCAGAACCTGATAATTA 300

Qy     301 TGGAGCATTCTACACTGACAGTTCTTTGAGACAAATTTCCTCTTGGCATTTACACTGTGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGGAGCATTCTACACTGACAGTTCTTTGAGACAAATTTCCTCTTGGCATTTACACTGTGG 360

Qy     361 CTTTAGCTTTCAAACCAGAGGTTCCTCTTACCCAGCAAAAAATGGCTTTCCCGAAGATGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTTTAGCTTTCAAACCAGAGGTTCCTCTTACCCAGCAAAAAATGGCTTTCCCGAAGATGA 420

Qy     421 GATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGTTTGTATTTTAGCATGTACA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGTTTGTATTTTAGCATGTACA 480

Qy     481 GTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAAGACGGGAACTTCCTTAAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAAGACGGGAACTTCCTTAAGC 540

Qy     541 TCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCC 600

Qy     601 ACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGA 660

Qy     661 AGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGAAACGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGAAACGA 720

Qy     721 AAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACG 780

Qy     781 TCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTC 840

Qy     841 CTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGA 900

Qy     901 CTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCA 960

Qy     961 ATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGT 1020

Qy    1021 GGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGA 1080

Qy    1081 GTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGG 1140

Qy    1141 GGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTT 1200

Qy    1201 TTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCA 1260

Qy    1261 TCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATTCCCGGGGGGGAAGATT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATTCCCGGGGGGGAAGATT 1320

Qy    1321 GTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGACAAGTTTCAGATAACCCGTGGG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGACAAGTTTCAGATAACCCGTGGG 1380

Qy    1381 GATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAGAGGATGCTGTTCTTCAGTGCT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAGAGGATGCTGTTCTTCAGTGCT 1440

Qy    1441 GAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTGATTAGTTCCCACTTGGTGCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTGATTAGTTCCCACTTGGTGCCC 1500

Qy    1501 CAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCTTCATACTAAGTGTTTGACATA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCTTCATACTAAGTGTTTGACATA 1560

Qy    1561 CACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGACATGATGGGTGTTACCATCCT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGACATGATGGGTGTTACCATCCT 1620

Qy    1621 AATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACAGAAACTTGTCCCAAGGCTCAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACAGAAACTTGTCCCAAGGCTCAC 1680

Qy    1681 AGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACTGTCAGAATCTGGTGGGCAGCC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 AGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACTGTCAGAATCTGGTGGGCAGCC 1740

Qy    1741 CCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGAGGGGACACTGATGATGAGGTC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
b    1741 CCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGAGGGGACACTGATGATGAGGTC 1800
y    1801 TCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCCACCTCAGCTCCCAGTCCTTGT 1860
b    1801 TCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCCACCTCAGCTCCCAGTCCTTGT 1860
y    1861 GCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCCAGCTCCTTCATTCACAGGGCT 1920
Db   1861 GCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCCAGCTCCTTCATTCACAGGGCT 1920
Qy   1921 GGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTACTCAACACAACCCAAGTTTCA 1980
Db   1921 GGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTACTCAACACAACCCAAGTTTCA 1980
Qy   1981 TCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGCTGAGGAACACCCAGCTCTGGA 2040
Db   1981 TCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGCTGAGGAACACCCAGCTCTGGA 2040
Qy   2041 CAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCACTGCATTCTGACCTGCTGTAC 2100
Db   2041 CAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCACTGCATTCTGACCTGCTGTAC 2100
Qy   2101 AGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACATCCCTTCTGAGCCATCTGCTG 2160
Db   2101 AGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACATCCCTTCTGAGCCATCTGCTG 2160
Qy   2161 CTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGATCAATGGGGACCAGCCACTGC 2220
Db   2161 CTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGATCAATGGGGACCAGCCACTGC 2220
Qy   2221 CCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACTCAGATGGGGTT 2280
Db   2221 CCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACTCAGATGGGGTT 2280
Qy   2281 CATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAA 2340
Db   2281 CATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAA 2340
Qy   2341 GGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCC 2400
Db   2341 GGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCC 2400
Qy   2401 ACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAAC 2460
Db   2401 ACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAAC 2460
Qy   2461 TGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTT 2520
Db   2461 TGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTT 2520
Qy   2521 TGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATT 2580
Db   2521 TGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATT 2580
Qy   2581 CAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCT 2640
Db   2581 CAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCT 2640
Qy   2641 TTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGGCTG 2700
Db   2641 TTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGGCTG 2700
Qy   2701 TGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2760
Db   2701 TGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2760
Qy   2761 AAAAAAAAAAAAAAA 2775
Db   2761 AAAAAAAAAAAAAAA 2775
```

RESULT 2
AB020337

```
LOCUS       AB020337                2762 bp    mRNA    linear   PRI 14-APR-2000
DEFINITION  Homo sapiens mRNA for UDP-Gal:GlcNAc beta1,3-galactosyltransferase
            5, complete cds.
ACCESSION   AB020337
VERSION     AB020337.1  GI:4835502
KEYWORDS    UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5.
SOURCE      Homo sapiens Adenocarcinoma cell_line:Colo 205 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isshiki,S., Togayachi,A., Kudo,T., Nishihara,S., Watanabe,M.,
            Kubota,T., Kitajima,M., Shiraishi,N., Sasaki,K., Andoh,T. and
            Narimatsu,H.
  TITLE     Cloning, expression, and characterization of a novel
            UDP-galactose:beta-N-acetylglucosamine
            beta1,3-galactosyltransferase (beta3Gal-T5) responsible for
            synthesis of type 1 chain in colorectal and pancreatic epithelia
            and tumor cells derived therefrom
  JOURNAL   J. Biol. Chem. 274 (18), 12499-12507 (1999)
  MEDLINE   99230269
REFERENCE   2  (bases 1 to 2762)
  AUTHORS   Isshiki,S., Togayachi,A. and Narimatsu,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-1998) Hisashi Narimatsu, Soka University,
            Institute of Life Science; 1-236, Tangi-cho, Hachioji, Tokyo
            192-8577, Japan (E-mail:sisshiki@po.iijnet.or.jp,
            Tel:81-426-91-9466, Fax:81-426-91-9315)
FEATURES             Location/Qualifiers
     source          1. .2762
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /cell_line="Colo 205"
                     /cell_type="Adenocarcinoma"
     gene            1. .2762
                     /gene="beta1,3-GalT 5"
     exon            1. .273
                     /gene="beta1,3-GalT 5"
                     /number=1
     exon            274. .433
                     /gene="beta1,3-GalT 5"
                     /number=3
     exon            434. .2762
                     /gene="beta1,3-GalT 5"
                     /number=4
     CDS             434. .1366
                     /gene="beta1,3-GalT 5"
                     /codon_start=1
                     /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
                     /protein_id="BAA77664.1"
                     /db_xref="GI:4835503"
                     /translation="MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNF
                     LKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSS
                     AAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
                     INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                     YVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                     LFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV"
     polyA_site      2762
                     /gene="beta1,3-GalT 5"
                     /note="45 a nucleotides"
BASE COUNT      651 a    703 c    679 g    729 t
ORIGIN

  Query Match           98.4%;  Score 2730;  DB 9;  Length 2762;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2730;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCA 60
Db      33 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCA 92
```

```
1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
90766 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 90825

1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
90826 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 90885

1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1299
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
90886 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 90945

1300 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 1359
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
90946 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 91005

1360 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 1419
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91006 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 91065

1420 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 1479
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91066 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 91125

1480 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 1539
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91126 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 91185

1540 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 1599
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91186 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 91245

1600 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 1659
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91246 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 91305

1660 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 1719
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91306 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 91365

1720 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 1779
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91366 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 91425

1780 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 1839
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91426 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 91485

1840 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 1899
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91486 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 91545

1900 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 1959
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91546 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 91605

1960 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 2019
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91606 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 91665

2020 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 2079
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91666 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 91725

2080 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 2139
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91726 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 91785

2140 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 2199
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
91786 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 91845
```

```
Qy   2200 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 2259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91846 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 91905

Qy   2260 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 2319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91906 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 91965

Qy   2320 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 2379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91966 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 92025

Qy   2380 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 2439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  92026 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 92085

Qy   2440 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 2499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  92086 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 92145

Qy   2500 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 2559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  92146 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 92205

Qy   2560 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 2619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  92206 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 92265

Qy   2620 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 2679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  92266 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 92325

Qy   2680 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAA 2739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  92326 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTACTTGAACT 92385

Qy   2740 AAAAAAAA 2747
          |  ||||
Db  92386 CAGGAAAA 92393
```

```
RESULT 4
HS21C080
LOCUS       HS21C080               340000 bp    DNA     linear   PRI 24-MAY-2000
DEFINITION  Homo sapiens chromosome 21 segment HS21C080.
ACCESSION   AL163280 AP001735 BA000005
VERSION     AL163280.2  GI:7717369
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
```

```
COMMENT     The Chromosome 21 Mapping and Sequencing Consortium consists of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: sakaki@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Dept. of Molecular Biology, *
            Tokyo 160-8582, Japan,
            * e.mail: shimizu@dmb-med.keio.ac.jp
            * URL: http://adenine.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/.
FEATURES              Location/Qualifiers
     source           1. .340000
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q22.3"
     source           <1. .125946
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q22.3"
                      /clone="P160L9, 5' partial"
                      /clone_lib="RPCI1,3-5 PAC library"
                      /note="Accession No. AF121897"
     source           73410. .243533
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q22.3"
                      /clone="P70I24"
                      /clone_lib="RPCI1,3-5 PAC library"
                      /note="Accession No. AF064860"
     source           224137. .>340000
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q22.3"
                      /clone="P206A10, 3' partial"
                      /clone_lib="RPCI1,3-5 PAC library"
                      /note="Accession No. AF121782"
     source           286628. .>340000
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q22.3"
                      /clone="BAC-291B3, 3' partial"
                      /clone_lib="BAC library"
                      /note="Accession No. AF064857"
     repeat_region    1088. .1566
                      /note="L1M4"
                      /rpt_family="LINE/L1"
                      /rpt_type=DISPERSED
     repeat_region    2398. .2495
                      /note="(TCTA)n"
                      /rpt_family="Simple_repeat"
                      /rpt_type=TANDEM
     repeat_region    complement(2496. .2878)
                      /note="THE1C"
                      /rpt_family="LTR/MaLR"
                      /rpt_type=DISPERSED
     repeat_region    2879. .2931
                      /note="(TCTA)n"
                      /rpt_family="Simple_repeat"
                      /rpt_type=TANDEM
     repeat_region    2932. .3004
                      /note="(TCCA)n"
                      /rpt_family="Simple_repeat"
                      /rpt_type=TANDEM
     repeat_region    3005. .3074
                      /note="(TCCA)n"
                      /rpt_family="Simple_repeat"
                      /rpt_type=TANDEM
     gene             3473. .19157
                      /gene="SH3BGR"
     mRNA             join(<3473. .3565,12649. .12678,15331. .15460,18739. .19157)
                      /gene="SH3BGR"
     CDS              join(3473. .3565,12649. .12678,15331. .15426)
                      /partial
                      /gene="SH3BGR"
                      /note="Accession No. X93498"
                      /codon_start=1
                      /product="21-Glutamic Acid Rich protein 21-GARP"
                      /protein_id="CAB90445.1"
                      /db_xref="GI:7717370"
                      /translation="GSEKAEEGGETEAQKEGSEDVGNLPEAQEKNEEEGETATEETEE
                      IAMEGAEGEAEEEEETAEGEEPGEDEDS"
     exon             3473. .3565
                      /gene="SH3BGR"
                      /number=4
     repeat_region    5542. .5829
                      /note="AluSx"
                      /rpt_family="SINE/Alu"
                      /rpt_type=DISPERSED
     repeat_region    complement(6887. .7067)
                      /note="L1MD3"
                      /rpt_family="LINE/L1"
                      /rpt_type=DISPERSED
     repeat_region    complement(8482. .8762)
                      /note="AluJo"
                      /rpt_family="SINE/Alu"
                      /rpt_type=DISPERSED
     repeat_region    complement(8865. .8964)
                      /note="L2"
                      /rpt_family="LINE/L2"
                      /rpt_type=DISPERSED
     repeat_region    9131. .9424
                      /note="AluSq"
                      /rpt_family="SINE/Alu"
                      /rpt_type=DISPERSED
     repeat_region    complement(9723. .9882)
                      /note="L2"
                      /rpt_family="LINE/L2"
                      /rpt_type=DISPERSED
     repeat_region    complement(9989. .10077)
                      /note="L2"
                      /rpt_family="LINE/L2"
                      /rpt_type=DISPERSED
     repeat_region    complement(10078. .10429)
                      /note="THE1B"
                      /rpt_family="LTR/MaLR"
                      /rpt_type=DISPERSED
     repeat_region    complement(10430. .10651)
                      /note="L2"
                      /rpt_family="LINE/L2"
                      /rpt_type=DISPERSED
     repeat_region    10980. .11000
                      /note="(TTTA)n"
                      /rpt_family="Simple_repeat"
                      /rpt_type=TANDEM
     repeat_region    complement(11002. .11282)
                      /note="AluSp"
```

```
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(11315. .11556)
                     /note="AluJo"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   11702. .11753
                     /note="MIR"
                     /rpt_family="SINE/MIR"
                     /rpt_type=DISPERSED
     exon            12649. .12678
                     /gene="SH3BGR"
                     /number=5
     repeat_region   complement(13115. .13425)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(13598. .13892)
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   14013. .14131
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"

  Query Match             84.0%;  Score 2332;  DB 9;  Length 340000;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 2338;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163454 AGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 163513

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163514 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 163573

Qy     520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163574 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 163633

Qy     580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163634 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 163693

Qy     640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163694 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 163753

Qy     700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163754 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 163813

Qy     760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163814 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 163873

Qy     820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163874 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 163933

Qy     880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163934 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 163993

Qy     940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163994 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 164053

Qy    1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164054 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 164113

Qy    1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164114 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 164173

Qy    1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164174 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 164233

Qy    1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164234 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 164293

Qy    1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164294 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 164353

Qy    1300 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 1359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164354 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 164413

Qy    1360 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 1419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164414 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 164473

Qy    1420 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 1479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164474 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 164533

Qy    1480 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 1539
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164534 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 164593

Qy    1540 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 1599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164594 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 164653

Qy    1600 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 1659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164654 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 164713

Qy    1660 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 1719
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164714 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 164773

Qy    1720 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 1779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164774 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 164833

Qy    1780 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 1839
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164834 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 164893

Qy    1840 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 1899
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164894 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 164953

Qy    1900 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 1959
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  164954 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 165013

Qy    1960 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 2019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  165014 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 165073

Qy    2020 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 2079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  165074 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 165133

Qy    2080 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 2139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  165134 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 165193

Qy    2140 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 2199
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165194 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 165253

Qy   2200 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 2259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165254 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 165313

Qy   2260 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 2319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165314 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 165373

Qy   2320 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 2379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165374 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 165433

Qy   2380 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 2439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165434 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 165493

Qy   2440 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 2499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165494 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 165553

Qy   2500 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 2559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165554 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 165613

Qy   2560 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 2619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165614 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 165673

Qy   2620 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 2679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165674 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 165733

Qy   2680 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAA 2739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||    ||
Db 165734 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTACTTGAACT 165793

Qy   2740 AAAAAAAA 2747
          |  ||||
Db 165794 CAGGAAAA 165801


RESULT 5
E38420
LOCUS       E38420                 10562 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Novel polypeptide.
ACCESSION   E38420
VERSION     E38420.1  GI:18626994
KEYWORDS    JP 2000245464-A/2.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 10562)
  AUTHORS   Narimatsu,H., Isshiki,S., Togayauchi,A. and Sasaki,K.
  TITLE     Novel polypeptide
  JOURNAL   Patent: JP 2000245464-A 2 12-SEP-2000;
            KYOWA HAKKO KOGYO CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 2000245464-A/2
            PD   12-SEP-2000
            PF   25-FEB-1999 JP 1999047571
            PR
            PI   HISASHI NARIMATSU,SOICHIRO ISSHIKI,AKIRA TOGAYAUCHI, PI
            KATSUTOSHI SASAKI
            PC   C12N15/09,A01K67/027,C12N1/21,C12N5/10,C12N9/10,C12P19/00, PC
             C12P21/02,
            PC   C12P21/08,C12Q1/68,G01N33/53//(C12N1/21,C12R1:185),(C12N5/10,
            PC   C12R1:91),
            PC   (C12P21/02,C12R1:185),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
            PC   (C12N5/00,C12R1:91)
            CC
            FH   Key             Location/Qualifiers
            FT   promoter        (1). .(5000)
            FT   exon            (5001). .(5140)
            FT   exon            (5001). .(5273)
            FT   exon            (5459). .(5567)
            FT   exon            (7427). .(7586)
            FT   exon            (8234). .(10562).
FEATURES             Location/Qualifiers
     source          1. .10562
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     2610 a    2415 c    2574 g    2963 t
ORIGIN

  Query Match             83.9%;  Score 2329.4;  DB 6;  Length 10562;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8232 AGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 8291

Qy    460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8292 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 8351

Qy    520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8352 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 8411

Qy    580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8412 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 8471

Qy    640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8472 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 8531

Qy    700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8532 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 8591

Qy    760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8592 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 8651

Qy    820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8652 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 8711

Qy    880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8712 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 8771

Qy    940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8772 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 8831

Qy   1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8832 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 8891

Qy   1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8892 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 8951

Qy   1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8952 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 9011

Qy   1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    9012 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 9071

Qy    1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9072 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 9131

Qy    1300 AGAATTCCCGGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 1359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9132 AGAATTCCCGGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 9191

Qy    1360 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 1419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9192 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 9251

Qy    1420 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 1479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9252 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 9311

Qy    1480 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 1539
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9312 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 9371

Qy    1540 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 1599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9372 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 9431

Qy    1600 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 1659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9432 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 9491

Qy    1660 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 1719
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9492 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 9551

Qy    1720 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 1779
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9552 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 9611

Qy    1780 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 1839
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9612 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 9671

Qy    1840 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 1899
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9672 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 9731

Qy    1900 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 1959
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9732 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 9791

Qy    1960 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 2019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9792 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 9851

Qy    2020 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 2079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9852 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 9911

Qy    2080 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 2139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9912 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 9971

Qy    2140 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 2199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9972 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 10031

Qy    2200 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 2259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10032 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 10091

Qy    2260 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 2319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10092 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 10151
```

```
Qy    2320 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 2379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10152 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 10211

Qy    2380 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 2439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10212 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 10271

Qy    2440 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 2499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10272 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 10331

Qy    2500 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 2559
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10332 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 10391

Qy    2560 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 2619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10392 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 10451

Qy    2620 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 2679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10452 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 10511

Qy    2680 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTT 2730
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10512 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTT 10562
```

RESULT 6
AF145784
LOCUS AF145784 933 bp DNA linear PRI 14-NOV-1999
DEFINITION Homo sapiens beta1,3 galactosyltransferase-V (B3GALT5) gene, complete cds.
ACCESSION AF145784
VERSION AF145784.1 GI:6409192
KEYWORDS .
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 933)
AUTHORS Zhou,D., Berger,E.G. and Hennet,T.
TITLE Molecular cloning of a human UDP-galactose:GlcNAcbeta1,3GalNAc beta1, 3 galactosyltransferase gene encoding an O-linked core3-elongation enzyme
JOURNAL Eur. J. Biochem. 263 (2), 571-576 (1999)
MEDLINE 99337698
PUBMED 10406968
REFERENCE 2 (bases 1 to 933)
AUTHORS Zhou,D. and Hennet,T.
TITLE Direct Submission
JOURNAL Submitted (26-APR-1999) Physiology, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland

```
FEATURES             Location/Qualifiers
     source          1. .933
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22"
     gene            1. .933
                     /gene="B3GALT5"
     CDS             1. .933
                     /gene="B3GALT5"
                     /note="glycosyltransferase"
                     /codon_start=1
                     /product="beta1,3 galactosyltransferase-V"
                     /protein_id="AAF07880.1"
                     /db_xref="GI:6409193"
                     /translation="MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNF
                     LKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSS
                     AAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
```

```
                    INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                    YVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                    LFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV"
BASE COUNT      227 a    232 c    241 g    233 t
ORIGIN

  Query Match             33.6%;  Score 933;  DB 9;  Length 933;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-236;
  Matches  933;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     402 ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGT 461
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGT 60

Qy     462 TTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 521
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 120

Qy     522 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 581
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 180

Qy     582 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 641
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 240

Qy     642 GGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 701
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 300

Qy     702 AGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 761
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 360

Qy     762 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 821
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 420

Qy     822 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTC 881
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTC 480

Qy     882 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 941
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 540

Qy     942 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 1001
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 600

Qy    1002 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 1061
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 660

Qy    1062 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 1121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 720

Qy    1122 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 1181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 780

Qy    1182 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 1241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 840

Qy    1242 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 1301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 900

Qy    1302 AATTCCCGGGGGGAAGATTGTCCGCCTGTCTGA 1334
           ||||||||||||||||||||||||||||||||
Db     901 AATTCCCGGGGGGAAGATTGTCCGCCTGTCTGA 933
```

```
RESULT 7
HSA6078
LOCUS       HSA6078                 933 bp    DNA     linear   PRI 11-MAY-2000
DEFINITION  Homo sapiens beta3gal-T6 gene.
ACCESSION   AJ006078
VERSION     AJ006078.1  GI:7799922
KEYWORDS    beta-1,3-galactosyltransferase; beta3gal-T6 gene.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 933)
  AUTHORS   Amado,M., Carneiro,F. and Clausen,H.
  TITLE     Cloning and expression of two beta-1,3-galactosyltransferases:
            beta3gal-T5 and beta3gal-T6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 933)
  AUTHORS   Amado,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) Amado M., Department of Oral Diagnostics,
            Royal dental School, Norre Alle 20, 2200 Copenhagen, DENMARK
FEATURES             Location/Qualifiers
     source          1. .933
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .933
                     /gene="beta3gal-T6"
     CDS             1. .933
                     /gene="beta3gal-T6"
                     /codon_start=1
                     /product="beta-1,3-galactosyltransferase"
                     /protein_id="CAB91547.1"
                     /db_xref="GI:7799923"
                     /translation="MAFPKMRLMYICLLVLGALCLYYSMNSLNPFKEQSFVYKKDGNF
                     LKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSS
                     AAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
                     INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                     YVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                     LFRRIVACHFIKPRTLLDYWQALENSRGEDCPPV"
BASE COUNT      229 a    234 c    240 g    229 t      1 others
ORIGIN

  Query Match             33.3%;  Score 924;  DB 9;  Length 933;
  Best Local Similarity   99.4%;  Pred. No. 5.4e-234;
  Matches  927;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy     402 ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGT 461
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1 ATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCTTCTGGTTCTGGGGGCTCTTTGT 60

Qy     462 TTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 521
           ||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTGTATTATAGCATGAACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 120

Qy     522 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 581
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 180

Qy     582 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 641
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 240

Qy     642 GGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 701
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 300

Qy     702 AGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 761
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGCAGTGCAGCGGAAACAAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 360

Qy     762 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 821
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 420

Qy     822 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTC 881
                ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     421 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCNGACATGTTC 480

Qy     882 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 941
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 540

Qy     942 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 1001
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 600

Qy    1002 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 1061
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 660

Qy    1062 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 1121
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 720

Qy    1122 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 1181
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 780

Qy    1182 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 1241
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 840

Qy    1242 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 1301
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 900

Qy    1302 AATTCCCGGGGGGAAGATTGTCCGCCTGTCTGA 1334
                ||||||||||||||||||||||||||||||||
Db     901 AATTCCCGGGGGGAAGATTGTCCGCCTGTCTGA 933


RESULT 8
AB041416
LOCUS       AB041416                1576 bp    DNA     linear   PRI 17-OCT-2000
DEFINITION  Homo sapiens beta1,3-GalT 5 gene for UDP-Gal:GlcNAc
            beta1,3-galactosyltransferase 5, partial cds.
ACCESSION   AB041416
VERSION     AB041416.1  GI:7593026
KEYWORDS    .
SOURCE      Homo sapiens (isolate:#056) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1576)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Silver Project
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 1576)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Naruya Saitou, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
            Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
            URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
            Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1576
                     /organism="Homo sapiens"
                     /isolate="#056"
                     /db_xref="taxon:9606"
                     /note="human sequence used for primer design based on Acc#
                     AB020337"
     exon            <1. .27
                     /number=3
     intron          28. .674
                     /number=3
     gene            675. .1576
                     /gene="beta1,3-GalT 5"
     CDS             675. .>1576
                     /gene="beta1,3-GalT 5"
                     /codon_start=1
                     /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
                     /protein_id="BAA94501.1"
                     /db_xref="GI:7593027"
                     /translation="MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNF
                     LKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSS
                     AAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
                     INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                     YVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                     LFRRIVACHFIKPRTLLDYWQALE"
     exon            675. .>1576
                     /gene="beta1,3-GalT 5"
                     /number=4
BASE COUNT      415 a    373 c    387 g    401 t
ORIGIN

  Query Match             32.4%;  Score 899.2;  DB 9;  Length 1576;
  Best Local Similarity   99.7%;  Pred. No. 2.3e-227;
  Matches  901;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 AGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 732

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     733 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 792

Qy     520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     793 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 852

Qy     580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     853 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 912

Qy     640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     913 GGGGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 972

Qy     700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     973 CCAGCAGTGCAGCGGAAACAAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 1032

Qy     760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1033 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 1092

Qy     820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1093 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 1152

Qy     880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1153 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 1212

Qy     940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1213 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 1272

Qy    1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1273 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1332

Qy    1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1333 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1392
```

```
Qy    1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1393 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1452

Qy    1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1453 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1512

Qy    1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1513 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1572

Qy    1300 AGAA 1303
           ||||
Db    1573 AGAA 1576
```

RESULT 9
AB041415
LOCUS AB041415 1570 bp DNA linear PRI 13-APR-2000
DEFINITION Pan paniscus beta1,3-GalT 5 gene for UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5, partial cds.
ACCESSION AB041415
VERSION AB041415.1 GI:7593024
KEYWORDS .
SOURCE Pan paniscus (isolate:bonobo-05) DNA.
ORGANISM Pan paniscus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 1570)
AUTHORS Liu,Y. and Saitou,N.
TITLE Silver Project
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 1570)
AUTHORS Liu,Y. and Saitou,N.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima, Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp, URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790, Fax:81-559-81-6789)

FEATURES Location/Qualifiers

```
     source          1. .1570
                     /organism="Pan paniscus"
                     /isolate="bonobo-05"
                     /db_xref="taxon:9597"
                     /note="human sequence used for primer design based on Acc#
                     AB020337"
     exon            <1. .20
                     /number=3
     intron          21. .667
                     /number=3
     gene            668. .1570
                     /gene="beta1,3-GalT 5"
     CDS             668. .>1570
                     /gene="beta1,3-GalT 5"
                     /codon_start=1
                     /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
                     /protein_id="BAA94500.1"
                     /db_xref="GI:7593025"
                     /translation="MAFPKMRLMYVCLLVLGALCLYFSMYSLNLFKEQSFVYKKDGNF
                     LKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSS
                     AAETKEVDQESQRHGDIIQKDFLDGYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
                     INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                     YVFSGDVASQVYNVSESVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                     RFRRIVACHFIKPRTLLDYWQALEN"
     exon            668. .>1570
                     /gene="beta1,3-GalT 5"
                     /number=4
```

BASE COUNT 408 a 372 c 392 g 398 t
ORIGIN

```
  Query Match           31.8%;  Score 882.6;  DB 9;  Length 1570;
  Best Local Similarity 98.5%;  Pred. No. 5.8e-223;
  Matches  891;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     666 AGATGGCTTTCCCGAAGATGAGATTGATGTATGTTTGCCTTCTGGTTCTGGGGGCTCTTT 725

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           |||||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||
Db     726 GTTTGTATTTTAGCATGTACAGTCTGAATCTTTTCAAAGAACAGTCCTTTGTTTACAAGA 785

Qy     520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     786 AAGATGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 845

Qy     580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 905

Qy     640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     906 GGGGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 965

Qy     700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 1025

Qy     760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
           |||||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||
Db    1026 TCCAGAAGGATTTCCTGGACGGCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 1085

Qy     820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 1145

Qy     880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 1205

Qy     940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    1206 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGTAAGTGGTTTG 1265

Qy    1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1266 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCGCCATTCTGCTCCGGCACCGGCTACG 1325

Qy    1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
           ||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1326 TGTTTTCCGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCGAGAGCGTCCCATACATTA 1385

Qy    1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1386 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1445

Qy    1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
           |||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db    1446 TCCACTCCCAGCCGACCTTTTTCCCAGGGGGCTTACGCTTCTCCGTATGCCGCTTCAGGA 1505

Qy    1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1506 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1565

Qy    1300 AGAAT 1304
           |||||
Db    1566 AGAAT 1570
```

RESULT 10
AB041413
LOCUS AB041413 1565 bp DNA linear PRI 13-APR-2000
DEFINITION Homo sapiens beta1,3-GalT 5 gene for UDP-Gal:GlcNAc

```
          beta1,3-galactosyltransferase 5, partial cds.
ACCESSION   AB041413
VERSION     AB041413.1  GI:7593020
KEYWORDS    .
SOURCE      Homo sapiens (isolate:human-NR) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1565)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Silver Project
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 1565)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Naruya Saitou, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
            Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
            URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
            Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1565
                     /organism="Homo sapiens"
                     /isolate="human-NR"
                     /db_xref="taxon:9606"
                     /note="human sequence used for primer design based on Acc#
                     AB020337"
     exon            <1. .27
                     /number=3
     intron          28. .674
                     /number=3
     gene            675. .1565
                     /gene="beta1,3-GalT 5"
     CDS             675. .>1565
                     /gene="beta1,3-GalT 5"
                     /codon_start=1
                     /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
                     /protein_id="BAA94498.1"
                     /db_xref="GI:7593021"
                     /translation="MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDGNF
                     LKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERXVKGKQLKTFFLLGTTSS
                     AAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
                     INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                     YVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                     LFRRIVACHFIKPRTLWTTGR"
     exon            675. .>1565
                     /gene="beta1,3-GalT 5"
                     /number=4
BASE COUNT      410 a    370 c    385 g    398 t      2 others
ORIGIN

  Query Match           31.7%;  Score 879.6;  DB 9;  Length 1565;
  Best Local Similarity   99.6%;  Pred. No. 3.6e-222;
  Matches  890;  Conservative    2;  Mismatches    1;  Indels    1;  Gaps    1;

Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 AGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 732

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     733 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 792

Qy     520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     793 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 852

Qy     580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     853 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 912

Qy     640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
           ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db     913 GGGGGAAAGAGAGGAYGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 972

Qy     700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
           ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db     973 CCAGCAGTGCAGCGGAAACRAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 1032

Qy     760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1033 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 1092

Qy     820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1093 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 1152

Qy     880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1153 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 1212

Qy     940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1213 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 1272

Qy    1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1273 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1332

Qy    1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1333 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1392

Qy    1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1393 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1452

Qy    1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1453 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1512

Qy    1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGG 1293
           |||||||||||||||||||||||||||||||| ||||||||||||||||
Db    1513 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTC-TGGACTACTGGCAGG 1565


RESULT 11
AB041412
LOCUS       AB041412               1360 bp    DNA     linear   PRI 17-OCT-2000
DEFINITION  Gorilla gorilla beta1,3-GalT 5 gene for UDP-Gal:GlcNAc
            beta1,3-galactosyltransferase 5, partial cds.
ACCESSION   AB041412
VERSION     AB041412.1  GI:7593018
KEYWORDS    .
SOURCE      Gorilla gorilla (isolate:#085) DNA.
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 1360)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Silver Project
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 1360)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Naruya Saitou, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
            Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
            URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
            Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1360
                     /organism="Gorilla gorilla"
                     /isolate="#085"
                     /db_xref="taxon:9593"
                     /note="human sequence used for primer design based on Acc#
```

```
                AB020337"
     intron     <1. .466
                /number=3
     gene       467. .1360
                /gene="beta1,3-GalT 5"
     CDS        467. .>1360
                /gene="beta1,3-GalT 5"
                /codon_start=1
                /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
                /protein_id="BAA94497.1"
                /db_xref="GI:7593019"
                /translation="MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKKDVNF
                LKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSS
                AAETKEVDQESRRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
                INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                YVFSGDVASQVYNVSESVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                RFRRIVACHFIKPRTLLDYWQA"
     exon       467. .>1360
                /gene="beta1,3-GalT 5"
                /number=4
BASE COUNT      334 a    331 c    341 g    351 t      3 others
ORIGIN

  Query Match             31.4%;  Score 872.4;  DB 9;  Length 1360;
  Best Local Similarity   98.1%;  Pred. No. 2.9e-220;
  Matches  879;  Conservative    3;  Mismatches   14;  Indels    0;  Gaps    0;

Qy    400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 AGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 524

Qy    460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    525 GTTTGTATTTTAGCATGTACAGTCTGAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 584

Qy    520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    585 AAGACGTGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 644

Qy    580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    645 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCAGGCAGACGT 704

Qy    640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
          |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db    705 GGGGGAAAGAGAGGACAGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 764

Qy    700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    765 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCGGCGACACGGGGACATTA 824

Qy    760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
          |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db    825 TCCAGAAGGATTTCCTAGAYGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 884

Qy    820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    885 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 944

Qy    880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    945 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 1004

Qy    940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1005 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGTAAGTGGTTTG 1064

Qy   1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
          |||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||||
Db   1065 TCAGTAAATCTGAATATCCATGGGACAGGTACCCGCCATTCTGCTCCGGCACCGGCTACG 1124

Qy   1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
          ||||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1125 TGTTTTCCGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCGAGAGCGTCCCATACATTA 1184

Qy   1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1185 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1244

Qy   1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
          ||||||||||||||||:||||| |||||||||||||||||||:||||||| ||||||||
Db   1245 TCCACTCCCAGCCGACSTTTTTCCCAGGGGGCTTACGCTTCTCYGTATGCCGCTTCAGGA 1304

Qy   1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCT 1295
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1305 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCT 1360


RESULT 12
AB041414
LOCUS       AB041414               1566 bp    DNA     linear   PRI 13-APR-2000
DEFINITION  Pan troglodytes beta1,3-GalT 5 gene for UDP-Gal:GlcNAc
            beta1,3-galactosyltransferase 5, partial cds.
ACCESSION   AB041414
VERSION     AB041414.1  GI:7593022
KEYWORDS    .
SOURCE      Pan troglodytes (isolate:chimp-202) DNA.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 1566)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Silver Project
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 1566)
  AUTHORS   Liu,Y. and Saitou,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Naruya Saitou, National Institute of
            Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima,
            Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp,
            URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790,
            Fax:81-559-81-6789)
FEATURES             Location/Qualifiers
     source          1. .1566
                     /organism="Pan troglodytes"
                     /isolate="chimp-202"
                     /db_xref="taxon:9598"
                     /note="human sequence used for primer design based on Acc#
                     AB020337"
     exon            <1. .28
                     /number=3
     intron          29. .675
                     /number=3
     gene            675. .1566
                     /gene="beta1,3-GalT 5"
     CDS             675. .>1566
                     /gene="beta1,3-GalT 5"
                     /codon_start=1
                     /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
                     /protein_id="BAA94499.1"
                     /db_xref="GI:7593023"
                     /translation="MAFPKMRLMYVCLLVLGALCVYFSMYSLNLFKEQSFVYKKDGNF
                     LKLPDTDCRQTPPFLVLLVTSSHRQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSS
                     AAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMF
                     INVDYLTELLLKKNRTTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTG
                     YVFSGDVASQVYNVSESVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVC
                     RFRRIVACHFIKPRTLLDYWQ"
     exon            675. .>1566
                     /gene="beta1,3-GalT 5"
                     /number=4
BASE COUNT      404 a    374 c    394 g    393 t      1 others
ORIGIN

  Query Match             31.3%;  Score 869.6;  DB 9;  Length 1566;
  Best Local Similarity   98.2%;  Pred. No. 1.6e-219;
  Matches  878;  Conservative    1;  Mismatches   15;  Indels    0;  Gaps    0;
```

```
Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     673 AGATGGCTTTCCCGAAGATGAGATTGATGTATGTTTGCCTTCTGGTTCTGGGGGCTCTTT 732

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           || |||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||
Db     733 GTGTGTATTTTAGCATGTACAGTCTGAATCTTTTCAAAGAACAGTCCTTTGTTTACAAGA 792

Qy     520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
           ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db     793 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCARACACCTCCCTTCCTCG 852

Qy     580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     853 TCCTGCTGGTGACCTCATCCCACAGACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 912

Qy     640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     913 GGGGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 972

Qy     700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     973 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATCA 1032

Qy     760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1033 TCCAGAAGGATTTCCTGGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 1092

Qy     820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1093 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 1152

Qy     880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1153 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 1212

Qy     940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    1213 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGTAAGTGGTTTG 1272

Qy    1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1273 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCGCCATTCTGCTCCGGCACCGGCTACG 1332

Qy    1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
           ||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1333 TGTTTTCCGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCGAGAGCGTCCCATACATTA 1392

Qy    1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1393 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1452

Qy    1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
           |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Db    1453 TCCACTCCCAGCCGACCTTTTTCCCAGGGGGCTTACGCTTCTCCGTATGCCGCTTCAGGA 1512

Qy    1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGG 1293
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1513 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGG 1566
```

RESULT 13
AB041417
LOCUS AB041417 1579 bp DNA linear PRI 13-APR-2000
DEFINITION Pongo pygmaeus beta1,3-GalT 5 gene for UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5, partial cds.
ACCESSION AB041417
VERSION AB041417.1 GI:7593028
KEYWORDS .
SOURCE Pongo pygmaeus (isolate:oran-Po17) DNA.
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE 1 (bases 1 to 1579)
AUTHORS Liu,Y. and Saitou,N.
TITLE Silver Project
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 1579)
AUTHORS Liu,Y. and Saitou,N.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Naruya Saitou, National Institute of Genetics, Laboratory of Evolutionary Genetics; 1111 Yata, Mishima, Shizuoka 411-8540, Japan (E-mail:nsaitou@genes.nig.ac.jp, URL:http://sayer.lab.nig.ac.jp/~silver/, Tel:81-559-81-6790, Fax:81-559-81-6789)

```
FEATURES             Location/Qualifiers
     source          1. .1579
                     /organism="Pongo pygmaeus"
                     /isolate="oran-Po17"
                     /db_xref="taxon:9600"
                     /note="human sequence used for primer design based on Acc#
                     AB020337"
     exon            <1. .26
                     /number=3
     intron          27. .673
                     /number=3
     gene            674. .1579
                     /gene="beta1,3-GalT 5"
     CDS             674. .>1579
                     /gene="beta1,3-GalT 5"
                     /codon_start=1
                     /product="UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5"
                     /protein_id="BAA94502.1"
                     /db_xref="GI:7593029"
                     /translation="MAFPKMRLIYICLLVLGALCLYFSMYNLNPFKEQSFVYKKEDGN
                     FLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTS
                     SAAEMKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGMEWVHRFCPQAAFVMKTDSDM
                     FINVDYLTELLLKKNRTTRFFTGFLKLNELPIRQPFSKWFVSKSEYPWDRYPPFCSGT
                     AYVFSGDVASQVYNVSESVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPEGLRFSV
                     CRFRRIVACHFIKPRTLLDYWQALEN"
     exon            674. .>1579
                     /gene="beta1,3-GalT 5"
                     /number=4
BASE COUNT 421 a 371 c 383 g 404 t
ORIGIN
```

Query Match 30.6%; Score 850.4; DB 9; Length 1579;
Best Local Similarity 96.8%; Pred. No. 2.1e-214;
Matches 879; Conservative 0; Mismatches 26; Indels 3; Gaps 1;

```
Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     672 AGATGGCTTTCCCGAAGATGAGATTGATTTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 731

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           |||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db     732 GTTTGTATTTTAGCATGTACAATCTGAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 791

Qy     520 AA---GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCC 576
           ||   |||||||||||||||||||||||||||||||||||||||| || ||||||||||||
Db     792 AAGAGGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGACAAACACCTCCCTTCC 851

Qy     577 TCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGA 636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     852 TCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGA 911

Qy     637 CGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGA 696
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     912 CGTGGGGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGA 971

Qy     697 CCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACA 756
           | |||||||||||| || |||| ||||||||||||||||||||||||||||||||||||
Db     972 CTACCAGCAGTGCGGCAGAAATGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACA 1031

Qy     757 TTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCA 816
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |
Db    1032 TTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTAAAGACCATGATGGGTA 1091

Qy     817 TAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACA 876
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1092 TGGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACA 1151

Qy     877 TGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGT 936
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1152 TGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGT 1211

Qy     937 TTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGT 996
           |||||||||||||||||||||||||| |||||||||||||||||||||| |||||||
Db    1212 TTTTCACTGGCTTCTTGAAACTCAATGAGCTTCCCATCAGGCAGCCATTCAGTAAGTGGT 1271

Qy     997 TTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCT 1056
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| | ||
Db    1272 TTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCTGGCACTGCCT 1331

Qy    1057 ACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACA 1116
           |||||||||| |||||||||| |||||||||||||||||||  ||||||||||||||||
Db    1332 ACGTGTTTTCCGGCGACGTGGCAAGTCAGGTGTACAATGTCTCTGAGAGCGTCCCATACA 1391

Qy    1117 TTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGG 1176
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1392 TTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGG 1451

Qy    1177 AGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCA 1236
           |||||||||||||||||||||||||| |||||||||||||||||| |||| |||||
Db    1452 AGCTCCACTCCCAGCCGACCTTTTTTCCAGAGGGCTTACGCTTCTCCGTGTGCCGCTTCA 1511

Qy    1237 GGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTC 1296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1512 GGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTC 1571

Qy    1297 TAGAGAAT 1304
           ||||||||
Db    1572 TAGAGAAT 1579


RESULT 14
AC020851
LOCUS       AC020851               196900 bp    DNA     linear   HTG 15-JUL-2000
DEFINITION  Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33
            unordered pieces.
ACCESSION   AC020851
VERSION     AC020851.2  GI:9211212
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:6686457.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1437240
            Center clone name: RPCI-21_467L12
            -------
            Summary Statistics
            Consensus quality: 164618 bases at least Q40
            Consensus quality: 180887 bases at least Q30
            Consensus quality: 183777 bases at least Q20
            Estimated insert size: 147000; pulse field gel estimation
            Estimated insert size: 193700; sum-of-contigs estimation
            Quality coverage:  5.67 in Q20 bases; pulse field gel estimation
            Quality coverage:  4.3 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1541: contig of 1541 bp in length
            *     1542     1641: gap of unknown length
            *     1642     3492: contig of 1851 bp in length
            *     3493     3592: gap of unknown length
            *     3593     4738: contig of 1146 bp in length
            *     4739     4838: gap of unknown length
            *     4839     7248: contig of 2410 bp in length
            *     7249     7348: gap of unknown length
            *     7349     9178: contig of 1830 bp in length
            *     9179     9278: gap of unknown length
            *     9279    11493: contig of 2215 bp in length
            *    11494    11593: gap of unknown length
            *    11594    13607: contig of 2014 bp in length
            *    13608    13707: gap of unknown length
            *    13708    15946: contig of 2239 bp in length
            *    15947    16046: gap of unknown length
            *    16047    18392: contig of 2346 bp in length
            *    18393    18492: gap of unknown length
            *    18493    21231: contig of 2739 bp in length
            *    21232    21331: gap of unknown length
            *    21332    24662: contig of 3331 bp in length
            *    24663    24762: gap of unknown length
            *    24763    27140: contig of 2378 bp in length
            *    27141    27240: gap of unknown length
            *    27241    29950: contig of 2710 bp in length
            *    29951    30050: gap of unknown length
            *    30051    32245: contig of 2195 bp in length
            *    32246    32345: gap of unknown length
            *    32346    35688: contig of 3343 bp in length
            *    35689    35788: gap of unknown length
            *    35789    40387: contig of 4599 bp in length
            *    40388    40487: gap of unknown length
            *    40488    43580: contig of 3093 bp in length
            *    43581    43680: gap of unknown length
            *    43681    46437: contig of 2757 bp in length
            *    46438    46537: gap of unknown length
            *    46538    50184: contig of 3647 bp in length
            *    50185    50284: gap of unknown length
            *    50285    54572: contig of 4288 bp in length
            *    54573    54672: gap of unknown length
            *    54673    60227: contig of 5555 bp in length
            *    60228    60327: gap of unknown length
            *    60328    68024: contig of 7697 bp in length
            *    68025    68124: gap of unknown length
            *    68125    78610: contig of 10486 bp in length
            *    78611    78710: gap of unknown length
            *    78711    88151: contig of 9441 bp in length
            *    88152    88251: gap of unknown length
            *    88252    96257: contig of 8006 bp in length
            *    96258    96357: gap of unknown length
            *    96358   106227: contig of 9870 bp in length
            *   106228   106327: gap of unknown length
            *   106328   115795: contig of 9468 bp in length
            *   115796   115895: gap of unknown length
            *   115896   126363: contig of 10468 bp in length
            *   126364   126463: gap of unknown length
            *   126464   137189: contig of 10726 bp in length
            *   137190   137289: gap of unknown length
            *   137290   146378: contig of 9089 bp in length
```

```
          *   146379   146478: gap of unknown length
          *   146479   161686: contig of 15208 bp in length
          *   161687   161786: gap of unknown length
          *   161787   176149: contig of 14363 bp in length
          *   176150   176249: gap of unknown length
          *   176250   196900: contig of 20651 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196900
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP21-467L12"
                     /clone_lib="RPCI mouse PAC library 21"
BASE COUNT    51456 a  43823 c  44608 g  53666 t   3347 others
ORIGIN

  Query Match           19.5%;  Score 542.2;  DB 2;  Length 196900;
  Best Local Similarity   73.8%;  Pred. No. 2.9e-132;
  Matches  734;  Conservative    0;  Mismatches  248;  Indels   13;  Gaps    3;

Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||  |  |||||  ||  |  | ||    | | ||||  |  |||| || || |
Db 102456 AGATGGCTCACATGAAGACAAGGCTAGTTTACGCCTCCATTCTCATGATGGGCGCACTCT 102515

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           | ||||| || |||||| |         | |||||| |||||   | |||||||  ||||
Db 102516 GCTTGTACTTCAGCATGGA---------TTCTTTCAGAGAACTCCCGTTTGTTTTTAAGA 102566

Qy     520 AAG---ACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCC 576
           ||    |||||| |||||| || | || |||| |||||||| |||||  || || ||||
Db 102567 AAAGTCACGGGAAGTTCCTTCAGATTCCGGATATAGACTGCAAGCAGAAGCCGCCTTTCC 102626

Qy     577 TCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGA 636
           | || |||||||||||| ||||| ||||| ||| |||| |  ||||||||||||||  |||
Db 102627 TGGTGCTGCTGGTGACGTCATCTCACAAGCAGCTGGCCGCTCGCATGGCCATCCGCAAGA 102686

Qy     637 CGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGA 696
           ||||||| | |||||     ||| ||||  | ||| ||| ||| |||||||| |||||||
Db 102687 CGTGGGGTAGAGAGACATCTGTGCAGGGCCAACAGGTGAGGACCTTCTTCCTTCTGGGGA 102746

Qy     697 CCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACA 756
           || ||  |||  | | ||| | | | |       || |||||| |||  ||| | ||||
Db 102747 CCTCCGACAGCACCGAGGAGATGGACGCCACAACCCTGGAGAGCGAGCAGCACCGCGACA 102806

Qy     757 TTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCA 816
           |||||||||||||| |||   || | ||| | ||| |||||||||||||||||||||| |
Db 102807 TTATCCAGAAGGACTTCAAGGATGCCTACTTCAACCTGACCCTGAAGACCATGATGGGTA 102866

Qy     817 TAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACA 876
           | |||||||| | | |||||||||||| | || |  |||||||||| ||||| ||||
Db 102867 TGGAATGGGTCTACCACTTTTGTCCTCAGACAGCTTACGTGATGAAAACGGACTCTGACA 102926

Qy     877 TGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGT 936
           ||||  | ||||||| ||||||||| |||||||| || |||||||||| ||| |||||||
Db 102927 TGTTTGTGAATGTTGGCTATCTGACGGAACTGCTGCTAAAGAAAAACAAAACGACCAGGT 102986

Qy     937 TTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGT 996
           | ||||| |||| | | || | | | || ||||||||| |||||   |||| ||||||||
Db 102987 TCTTCACAGGCTACATAAAGCCCCACGACTTTCCCATCCGGCAGAAGTTCAACAAGTGGT 103046

Qy     997 TTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCT 1056
           |||| ||||| |  |||||||| |||||||||||||||| || |||||||| || || |
Db 103047 TTGTGAGTAAGTTCGAATATCCCTGGGACAGGTACCCACCTTTTTGCTCCGGTACTGGTT 103106

Qy    1057 ACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACA 1116
           | || |||||  |||||||||||  || || ||||| |||||  ||||||| || | ||
Db 103107 ATGTCTTTTCCAGCGACGTGGCGATCCAAGTATACAACGTCTCAGAGAGCGTTCCGTTCA 103166

Qy    1117 TTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGG 1176
           | || ||||| || |||||||| |||||||| | |  | ||| || ||| |   |||||
Db 103167 TCAAGCTGGAGGATGTGTTTGTTGGGCTCTGCTTGGCCAAGCTAAAGATCCGGCCGGAGG 103226

Qy    1177 AGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCA 1236
           |||| ||| || | | |||||| || || || || |||||||| ||||| |||| |||
Db 103227 AGCTGCACACCAAACAGACCTTCTTCCCTGGCGGTTTACGCTTTTCCGTGTGCCGCTTTC 103286

Qy    1237 GGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTC 1296
           ||  || ||||| ||||| || || ||||| | |   ||  |    |||||||| || |
Db 103287 AGAAAATTGTGGCATGCCATTTTATGAAGCCCCAGGACCTGCTCACTTACTGGCAAGCAC 103346

Qy    1297 TAGAGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCG 1356
           | ||||| ||    |   | || |||||  |||||||||||| ||  | ||||| |  |
Db 103347 TGGAGAACTCGAAAGAACAGGACTGTCCTGCTGTCTGAGGGGACCCTGGGGGCACCT-AG 103405

Qy    1357 GACAAGTTTCAGATAACCCGTGGGGATAGTTTTTG 1391
           |||||  ||| |||||||  ||| |||  ||| ||
Db 103406 GACAAACTTCTGATAACCTTTGGTGATGATTTGTG 103440


RESULT 15
AC120346/c
LOCUS       AC120346              175861 bp    DNA     linear   HTG 09-MAY-2002
DEFINITION  Mus musculus clone RP23-147E11, WORKING DRAFT SEQUENCE, 22 ordered
            pieces.
ACCESSION   AC120346
VERSION     AC120346.3  GI:20514894
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-147E11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
```

```
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 9, 2002 this sequence version replaced gi:20503171.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L20921
                Center clone name: 147_E_11
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 169879 bases at least Q40
                Consensus quality: 172188 bases at least Q30
                Consensus quality: 173133 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 173761; sum-of-contigs
                Quality coverage: 8.3 in Q20 bases; agarose-fp
                Quality coverage: 8.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    28835: contig of 28835 bp in length
            *    28836 28935: gap of      100 bp
            *    28936    29562: contig of 627 bp in length
            *    29563 29662: gap of      100 bp
            *    29663    30328: contig of 666 bp in length
            *    30329 30428: gap of      100 bp
            *    30429    32005: contig of 1577 bp in length
            *    32006 32105: gap of      100 bp
            *    32106    33128: contig of 1023 bp in length
            *    33129 33228: gap of      100 bp
            *    33229    34512: contig of 1284 bp in length
            *    34513 34612: gap of      100 bp
            *    34613    36079: contig of 1467 bp in length
            *    36080 36179: gap of      100 bp
            *    36180    37483: contig of 1304 bp in length
            *    37484 37583: gap of      100 bp
            *    37584    39166: contig of 1583 bp in length
            *    39167 39266: gap of      100 bp
            *    39267    41065: contig of 1799 bp in length
            *    41066 41165: gap of      100 bp
            *    41166    43203: contig of 2038 bp in length
            *    43204 43303: gap of      100 bp
            *    43304    47489: contig of 4186 bp in length
            *    47490 47589: gap of      100 bp
            *    47590    50604: contig of 3015 bp in length
            *    50605 50704: gap of      100 bp
            *    50705    56370: contig of 5666 bp in length
            *    56371 56470: gap of      100 bp
            *    56471    64079: contig of 7609 bp in length
            *    64080 64179: gap of      100 bp
            *    64180    72619: contig of 8440 bp in length
            *    72620 72719: gap of      100 bp
            *    72720    81566: contig of 8847 bp in length
            *    81567 81666: gap of      100 bp
            *    81667    96774: contig of 15108 bp in length
            *    96775 96874: gap of      100 bp
            *    96875   119647: contig of 22773 bp in length
            *   119648 119747: gap of      100 bp
            *   119748   147772: contig of 28025 bp in length
            *   147773 147872: gap of      100 bp
            *   147873   174557: contig of 26685 bp in length
            *   174558 174657: gap of      100 bp
            *   174658   175861: contig of 1204 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175861
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-147E11"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .28835
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28936. .29562
                     /note="assembly_fragment"
     misc_feature    29663. .30328
                     /note="assembly_fragment"
     misc_feature    30429. .32005
                     /note="assembly_fragment"
     misc_feature    32106. .33128
                     /note="assembly_fragment"
     misc_feature    33229. .34512
                     /note="assembly_fragment"
     misc_feature    34613. .36079
                     /note="assembly_fragment"
     misc_feature    36180. .37483
                     /note="assembly_fragment"
     misc_feature    37584. .39166
                     /note="assembly_fragment"
     misc_feature    39267. .41065
                     /note="assembly_fragment"
     misc_feature    41166. .43203
                     /note="assembly_fragment"
     misc_feature    43304. .47489
                     /note="assembly_fragment"
     misc_feature    47590. .50604
                     /note="assembly_fragment"
     misc_feature    50705. .56370
                     /note="assembly_fragment"
     misc_feature    56471. .64079
                     /note="assembly_fragment"
     misc_feature    64180. .72619
                     /note="assembly_fragment"
     misc_feature    72720. .81566
                     /note="assembly_fragment"
     misc_feature    81667. .96774
                     /note="assembly_fragment"
     misc_feature    96875. .119647
                     /note="assembly_fragment"
     misc_feature    119748. .147772
                     /note="assembly_fragment"
     misc_feature    147873. .174557
                     /note="assembly_fragment"
     misc_feature    174658. .175861
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48903 a  39540 c  39029 g  46286 t   2103 others
ORIGIN

  Query Match             19.3%;  Score 534.4;  DB 2;  Length 175861;
  Best Local Similarity   73.4%;  Pred. No. 3.4e-130;
  Matches  728;  Conservative    0;  Mismatches  251;  Indels   13;  Gaps    3;
```

```
Qy     400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
           | ||||||  |  |||||  ||  |  | ||    | | ||||  |  |||| || || |
Db   62875 AGATGGCTCACATGAAGACAAGGCTAGTTTACGCCTCCATTCTCATGATGGGCGCACTCT 62816

Qy     460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
           | ||||| || |||||| |          | |||||| |||||   | |||||||  ||||
Db   62815 GCTTGTACTTCAGCATGGA---------TTCTTTCAGAGAACTCCCGTTTGTTTTTAAGA 62765

Qy     520 AAG---ACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCC 576
           ||    ||||||| |||||| || | || |||| |||||||| |||||  || || ||||
Db   62764 AAAGTCACGGGAAGTTCCTTCAGATTCCGGATATAGACTGCAAGCAGAAGCCGCCTTTCC 62705

Qy     577 TCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGA 636
           | || ||||||||||| ||||| ||||| ||| |||| |  ||||||||||||||  |||
Db   62704 TGGTGCTGCTGGTGACGTCATCTCACAAGCAGCTGGCCGCTCGCATGGCCATCCGCAAGA 62645

Qy     637 CGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGA 696
           |||||||  | |||||     ||| ||||  | ||| ||| ||| |||||||| ||||||| 
Db   62644 CGTGGGGTAGAGAGACATCTGTGCAGGGCCAACAGGTGAGGACCTTCTTCCTTCTGGGGA 62585

Qy     697 CCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACA 756
           || ||  |||  | | ||| | | | |       || ||||||| |||  ||| | ||||
Db   62584 CCTCCGACAGCACCGAGGAGATGGACGCCACAACCCTGGAGAGCGAGCAGCACCGCGACA 62525

Qy     757 TTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCA 816
           |||||||||||| |||   || | ||| | ||| ||||||||||| ||||||||||| |
Db   62524 TTATCCAGAAGGACTTCAAGGATGCCTACTTCAACCTGACCCTGAAAACCATGATGGGTA 62465

Qy     817 TAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACA 876
           | ||||||||| | | |||||||||||||| | || |  |||||||||| ||||| ||||
Db   62464 TGGAATGGGTCTACCACTTTTGTCCTCAGACAGCTTACGTGATGAAAACGGACTCTGACA 62405

Qy     877 TGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGT 936
           ||||  | ||||||| |||||||| |||||||| || |||||||||| ||| |||||||
Db   62404 TGTTTGTGAATGTTGGCTATCTGACGGAACTGCTGCTAAAGAAAAACAAAACGACCAGGT 62345

Qy     937 TTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGT 996
           | ||||| |||| | | || | | | || |||||||||| |||||   |||| ||||||||
Db   62344 TCTTCACAGGCTACATAAAGCCCCACGACTTTCCCATCCGGCAGAAGTTCAACAAGTGGT 62285

Qy     997 TTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCT 1056
           |||| |||||  |  |||||||| |||||||||||||||||| || ||||| || || || |
Db   62284 TTGTGAGTAAGTTCGAATATCCCTGGGACAGGTACCCACCTTTTTGCTCTGGTACTGGTT 62225

Qy    1057 ACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACA 1116
           | || |||||  ||||||||||||  || || ||||| |||||  |||||| || | ||
Db   62224 ATGTCTTTTCCAGCGACGTGGCGATCCAAGTATACAACGTCTCAGAGAGCGTTCCGTTCA 62165

Qy    1117 TTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGG 1176
           | || ||||| || |||||||| ||||||||| | |  | ||| || ||| |   |||||
Db   62164 TCAAGCTGGAGGATGTGTTTGTTGGGCTCTGCTTGGCCAAGCTAAAGATCCGGCCGGAGG 62105

Qy    1177 AGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCA 1236
           |||| ||| || | | |||||| || || || || |||||||| ||||| |||| |||
Db   62104 AGCTGCACACCAAACAGACCTTCTTCCCTGGCGGTTTACGCTTTTCCGTGTGCCGCTTTC 62045

Qy    1237 GGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTC 1296
             ||  || ||||| ||||| || || ||||| | |   ||  |    |||||||| || |
Db   62044 AGAAAATTGTGGCATGCCATTTTATGAAGCCCCAGGACCTGCTCACTTACTGGCAAGCAC 61985

Qy    1297 TAGAGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCG 1356
           | ||||| ||    |  | || || ||  |||||||||||| ||  | ||| | |  |
Db   61984 TGGAGAACTCGAAAGAACAGGACTGCCCTGCTGTCTGAGGGGACCCTGGGGGCGCCT-AG 61926

Qy    1357 GACAAGTTTCAGATAACCCGTGGGGATAGTTT 1388
           |||||  ||| |||||||  ||| |||  |||
Db   61925 GACAAACTTCTGATAACCTTTGGTGATGATTT 61894


Search completed: April 11, 2003, 20:10:08
Job time : 6662.19 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:50:38 ; Search time 461.132 Seconds
(without alignments)
13552.085 Million cell updates/sec

Title: US-09-914-152-2
Perfect score: 2775
Sequence: 1 gtgaattcctctttctctgc.....aaaaaaaaaaaaaaaaaaaa 2775

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2775	100.0	2775	21	AAA93875	Human beta-1,3 gal
	2	2329.4	83.9	10562	21	AAA93876	Human beta3Gal-T5
	3	928.6	33.5	1011	21	AAA27959	Human beta3Gal-T5
	4	163.4	5.9	1739	15	AAQ67067	Beta-1,3-galactosy
	5	133.6	4.8	3518	19	AAV57905	Bovine butyrophili
	6	130.4	4.7	1773	20	AAX87193	Human Dendriac cDN
	7	130.4	4.7	1897	24	ABK51201	Human cDNA encodin
	8	130.4	4.7	2095	21	AAZ65022	Membrane-bound pro
	9	130.4	4.7	2095	22	AAS46005	Human DNA encoding
	10	130.4	4.7	2095	22	AAF92075	Human PRO1074 cDNA
	11	130.4	4.7	2095	22	AAF44168	Human PRO1074 (UNQ
	12	130.4	4.7	2168	22	AAH15711	Human cDNA sequenc
	13	130.4	4.7	2189	20	AAX87192	Human Dendriac cDN
	14	124.8	4.5	333	20	AAX51967	Human secreted pro
	15	124	4.5	2420	20	AAX35711	cDNA encoding a pr
	16	122.4	4.4	1266	20	AAX35710	cDNA encoding a pr
	17	119.8	4.3	847	21	AAC69087	Human secreted pro
c	18	115.8	4.2	9289	22	AAK90960	Human digestive sy
c	19	115.8	4.2	9289	22	AAS31995	Human liver associ
c	20	115.8	4.2	9289	24	ABN90350	Human liver antige
c	21	115.8	4.2	17865	22	AAK90961	Human digestive sy
c	22	115.8	4.2	17865	22	AAS31996	Human liver associ
c	23	115.8	4.2	17865	24	ABN90351	Human liver antige
	24	114.8	4.1	1134	21	AAA58791	DNA encoding the b
	25	111.6	4.0	671	22	AAK52784	Human polynucleoti
	26	111.6	4.0	827	22	AAK51800	Human polynucleoti
c	27	108.4	3.9	125439	24	ABQ88177	Human osteoblast d
	28	108.2	3.9	1134	22	AAD09356	Human beta-1,3-gal
	29	106	3.8	303	24	ABN24320	Human ORFX polynuc
	30	105	3.8	474	22	ABA59003	Human foetal liver
	31	105	3.8	474	22	ABA27841	Probe #6307 for ge
	32	105	3.8	474	22	AAK07173	Human brain expres
	33	105	3.8	474	22	AAK32922	Human bone marrow
	34	105	3.8	474	22	AAI16260	Probe #6193 for ge
	35	105	3.8	474	22	AAI38728	Probe #7414 used t
	36	105	3.8	474	24	ABS07727	Human genome-deriv
	37	103.8	3.7	37808	20	AAX02780	Vector pMVX-BG DNA
c	38	97.8	3.5	160271	22	AAF85750	Bipolar affective
c	39	97.8	3.5	160271	22	AAF85756	Human chromosome 1
c	40	97.8	3.5	160271	22	AAS04858	Human chromosome 1
c	41	97.8	3.5	160271	22	AAS04864	Human chromosome 1
c	42	97.8	3.5	160271	22	AAS06667	Human chromosome 1
c	43	97.8	3.5	160271	22	AAH40997	160kb fragment of
c	44	97.8	3.5	160271	22	AAH23764	Human chromosome 1
c	45	97.8	3.5	160271	22	AAF85116	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAA93875
ID AAA93875 standard; DNA; 2775 BP.
XX
AC AAA93875;
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta-1,3 galactose transferase encoding DNA.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; ds.
XX
OS Homo sapiens.
XX
PN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI; 2000-549409/50.
DR P-PSDB; AAB93875.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of

PT digestive system cancer -
XX
PS Claim 5; Page 99-102; 123pp; Japanese.
XX
CC This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAcbeta1-3Galbeta1-4Glc) to give Galbeta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents Beta-1,3 galactose
CC transferase encoding DNA.
XX
SQ Sequence 2775 BP; 681 A; 698 C; 669 G; 727 T; 0 other;

Query Match 100.0%; Score 2775; DB 21; Length 2775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCA 60

Qy      61 GAGCTGCAGGCTCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAGCTGCAGGCTCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTC 120

Qy     121 TCAGGCCTTTGGCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCAGGCCTTTGGCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTT 180

Qy     181 GGTCTTGGACTGAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGTCTTGGACTGAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCAC 240

Qy     241 GTGATTCCTGTCAGAATCACCATTTTTGGTAAACAAACCAAGCCCAGAACCTGATAATTA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTGATTCCTGTCAGAATCACCATTTTTGGTAAACAAACCAAGCCCAGAACCTGATAATTA 300

Qy     301 TGGAGCATTCTACACTGACAGTTCTTTGAGACAAATTTCCTCTTGGCATTTACACTGTGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGGAGCATTCTACACTGACAGTTCTTTGAGACAAATTTCCTCTTGGCATTTACACTGTGG 360

Qy     361 CTTTAGCTTTCAAACCAGAGGTTCCTCTTACCCAGCAAAAAATGGCTTTCCCGAAGATGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTTTAGCTTTCAAACCAGAGGTTCCTCTTACCCAGCAAAAAATGGCTTTCCCGAAGATGA 420

Qy     421 GATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGTTTGTATTTTAGCATGTACA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGTTTGTATTTTAGCATGTACA 480

Qy     481 GTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAAGACGGGAACTTCCTTAAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAAGACGGGAACTTCCTTAAGC 540

Qy     541 TCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCC 600

Qy     601 ACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGA 660

Qy     661 AGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGAAACGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGAAACGA 720

Qy     721 AAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACG 780

Qy     781 TCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTC 840

Qy     841 CTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGA 900

Qy     901 CTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCA 960

Qy     961 ATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGT 1020

Qy    1021 GGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGA 1080

Qy    1081 GTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGG 1140

Qy    1141 GGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTT 1200

Qy    1201 TTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCA 1260

Qy    1261 TCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATTCCCGGGGGGGAAGATT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATTCCCGGGGGGGAAGATT 1320

Qy    1321 GTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGACAAGTTTCAGATAACCCGTGGG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGACAAGTTTCAGATAACCCGTGGG 1380

Qy    1381 GATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAGAGGATGCTGTTCTTCAGTGCT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAGAGGATGCTGTTCTTCAGTGCT 1440

Qy    1441 GAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTGATTAGTTCCCACTTGGTGCCC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTGATTAGTTCCCACTTGGTGCCC 1500

Qy    1501 CAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCTTCATACTAAGTGTTTGACATA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCTTCATACTAAGTGTTTGACATA 1560

Qy    1561 CACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGACATGATGGGTGTTACCATCCT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGACATGATGGGTGTTACCATCCT 1620

Qy    1621 AATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACAGAAACTTGTCCCAAGGCTCAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACAGAAACTTGTCCCAAGGCTCAC 1680

Qy    1681 AGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACTGTCAGAATCTGGTGGGCAGCC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 AGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACTGTCAGAATCTGGTGGGCAGCC 1740

Qy    1741 CCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGAGGGGACACTGATGATGAGGTC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGAGGGGACACTGATGATGAGGTC 1800

Qy    1801 TCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCCACCTCAGCTCCCAGTCCTTGT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1801 TCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCCACCTCAGCTCCCAGTCCTTGT 1860

Qy   1861 GCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCCAGCTCCTTCATTCACAGGGCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCCAGCTCCTTCATTCACAGGGCT 1920

Qy   1921 GGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTACTCAACACAACCCAAGTTTCA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTACTCAACACAACCCAAGTTTCA 1980

Qy   1981 TCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGCTGAGGAACACCCAGCTCTGGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGCTGAGGAACACCCAGCTCTGGA 2040

Qy   2041 CAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCACTGCATTCTGACCTGCTGTAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCACTGCATTCTGACCTGCTGTAC 2100

Qy   2101 AGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACATCCCTTCTGAGCCATCTGCTG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 AGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACATCCCTTCTGAGCCATCTGCTG 2160

Qy   2161 CTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGATCAATGGGGACCAGCCACTGC 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGATCAATGGGGACCAGCCACTGC 2220

Qy   2221 CCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACTCAGATGGGGTT 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACTCAGATGGGGTT 2280

Qy   2281 CATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAA 2340

Qy   2341 GGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCC 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCC 2400

Qy   2401 ACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAAC 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 ACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAAC 2460

Qy   2461 TGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTT 2520

Qy   2521 TGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATT 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATT 2580

Qy   2581 CAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCT 2640

Qy   2641 TTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGGCTG 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGGCTG 2700

Qy   2701 TGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 TGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2760

Qy   2761 AAAAAAAAAAAAAAA 2775
          |||||||||||||||
Db   2761 AAAAAAAAAAAAAAA 2775


RESULT 2
AAA93876
ID   AAA93876 standard; DNA; 10562 BP.
XX
AC   AAA93876;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Human beta3Gal-T5 encoding DNA.
XX
KW   Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW   digestive system; beta3Gal-T5; ds.
XX
OS   Homo sapiens.
XX
PN   WO200050608-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-JP01070.
XX
PR   25-FEB-1999;   99JP-0047571.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Narimatsu H,  Isshiki S,  Togayachi A,  Sasaki K;
XX
DR   WPI; 2000-549409/50.
XX
PT   Beta-1,3 galactose transferase and DNA encoding it, useful for
PT   synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT   digestive system cancer -
XX
PS   Claim 31; Page 103-111; 123pp; Japanese.
XX
CC   This invention relates to a polypeptide (I) with beta-1,3 galactose
CC   transferase activity, or variants of (I) comprising amino acid additions,
CC   deletions and/or substitutions. Included in the invention is DNA encoding
CC   all or part of (I); expression vectors containing the DNA, host cells
CC   transformed by the vectors; a method for the preparation of the
CC   polypeptide by culture of the transformants or by expression in the milk
CC   of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC   galactose transferase protein transfers galactose by beta-1,3 bonding to
CC   N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC   GlcNAcbeta1-3Galbeta1-4Glc) to give Galbeta1-3GlcNAc. The protein and
CC   DNA encoding it are useful for the treatment and diagnosis of cancer of
CC   the digestive system. The present sequence represents a Beta3Gal-T5
CC   encoding DNA sequence.
XX
SQ   Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;

  Query Match             83.9%;  Score 2329.4;  DB 21;  Length 10562;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8232 AGATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 8291

Qy    460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8292 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 8351

Qy    520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8352 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 8411

Qy    580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8412 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 8471

Qy    640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8472 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 8531

Qy    700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8532 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 8591
```

```
Qy    760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8592 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 8651

Qy    820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8652 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 8711

Qy    880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8712 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 8771

Qy    940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8772 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 8831

Qy   1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8832 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 8891

Qy   1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8892 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 8951

Qy   1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8952 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 9011

Qy   1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9012 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 9071

Qy   1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9072 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 9131

Qy   1300 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 1359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9132 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGAC 9191

Qy   1360 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 1419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9192 AAGTTTCAGATAACCCGTGGGGATAGTTTTTGCTAGATTTTGGAAGAGGGGGCGGGACAG 9251

Qy   1420 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 1479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9252 AGGATGCTGTTCTTCAGTGCTGAAATCCACGCCAGAATGTCGGTGTTCATGAAGTCACTG 9311

Qy   1480 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 1539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9312 ATTAGTTCCCACTTGGTGCCCCAGGCAATAATAGGCCCGTCTCTTGGGCACGCACACTCT 9371

Qy   1540 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 1599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9372 TCATACTAAGTGTTTGACATACACCTGGATTTTTGCATTTCAGGGGTCAGTATCCTATGA 9431

Qy   1600 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 1659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9432 CATGATGGGTGTTACCATCCTAATTTTACAGGCAAGGACACAGCAGCTGCGAGAGGTACA 9491

Qy   1660 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 1719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9492 GAAACTTGTCCCAAGGCTCACAGCCAGTAGGCATAGGAGCGGGAATGAAAATCGAGCACT 9551

Qy   1720 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 1779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9552 GTCAGAATCTGGTGGGCAGCCCCTGACTTGAACCACTCCCACGTGCTGCCTCCCTTAGGA 9611

Qy   1780 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 1839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9612 GGGGACACTGATGATGAGGTCTCGGAGCCGGCATCCTTCCATCCCTGTCGAGTCCCCTCC 9671

Qy   1840 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 1899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9672 ACCTCAGCTCCCAGTCCTTGTGCTTTTTGGAGCTAAGCCTGGGATGACCAAATTCACCCC 9731

Qy   1900 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 1959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9732 AGCTCCTTCATTCACAGGGCTGGATGTAGCTGGGATTGAGTCCATGTTATCGGCTCGGTA 9791

Qy   1960 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 2019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9792 CTCAACACAACCCAAGTTTCATCCGAGGAAATGTCCCCGCAGTGGATGCAGCTCACATGC 9851

Qy   2020 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 2079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9852 TGAGGAACACCCAGCTCTGGACAGAGTTCTTATAAATGTATAAATTAGGCTCAGAAACCA 9911

Qy   2080 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 2139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9912 CTGCATTCTGACCTGCTGTACAGACTGCCCACACTGCTGACCTGCCTAGCGAGCAGGACA 9971

Qy   2140 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 2199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9972 TCCCTTCTGAGCCATCTGCTGCTCTCTCATTTCATCACCCCAACTGTCCCTTGTTTTTGA 10031

Qy   2200 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 2259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10032 TCAATGGGGACCAGCCACTGCCCCAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGA 10091

Qy   2260 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 2319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10092 CAGTTGAACTCAGATGGGGTTCATGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGA 10151

Qy   2320 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 2379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10152 GTCAAGTCAGGATGCTCTCAAGGACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTG 10211

Qy   2380 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 2439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10212 AAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGT 10271

Qy   2440 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 2499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10272 CTCATTTGGCCGAGGAACAACTGAACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCC 10331

Qy   2500 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 2559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10332 GCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGT 10391

Qy   2560 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 2619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10392 ACTTCGTTAGTTTTGCCCATTCAGCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCT 10451

Qy   2620 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 2679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10452 AATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAG 10511

Qy   2680 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTT 2730
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10512 AGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTT 10562
```

```
RESULT 3
AAA27959
ID   AAA27959 standard; DNA; 1011 BP.
XX
AC   AAA27959;
XX
DT   15-AUG-2000  (first entry)
XX
DE   Human beta3Gal-T5 gene sequence.
XX
KW   UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;
KW   Beta3Gal-T5; ss; human; chromosome 21q22.3; galactosylation;
KW   beta1,3-galactosyl glycosylated saccharide production; glycopeptide;
```

```
KW   glycoprotein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..1011
FT                   /*tag= a
FT                   /product= "Beta3Gal-T5"
FT   primer_bind     complement (79..98)
FT                   /*tag= b
FT   primer_bind     complement (150..170)
FT                   /*tag= c
FT   primer_bind     991..1011
FT                   /*tag= d
XX
PN   WO200029558-A1.
XX
PD   25-MAY-2000.
XX
PF   11-NOV-1999;   99WO-US26807.
XX
PR   13-NOV-1998;   98DK-0001483.
XX
PA   (CLAU/) CLAUSEN H.
XX
PI   Clausen H,  Amado M;
XX
DR   WPI; 2000-399728/34.
DR   P-PSDB; AAY94641.
XX
PT   Novel nucleic acid sequence encoding human
PT   UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
PT   useful for obtaining beta 1,3-galactosyl glycosylated saccharides and
PT   glycopeptides or glycoproteins -
XX
PS   Claim 7; Fig 1; 74pp; English.
XX
CC   The present invention relates to a nucleic acid sequence encoding
CC   UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
CC   (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at
CC   carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence
CC   represents the human beta3Gal-T5 gene sequence. The beta3Gal-T5 gene is
CC   located on human chromosome 21q22.3. Beta3Gal-T5 is a type II
CC   transmembrane glycoprotein. The invention also relates to the beta3Gal-T5
CC   protein sequence, a nucleic acid vector comprising the beta3Gal-T5
CC   nucleotide sequence, a host cell comprising the vector, and a method for
CC   the production of the beta3Gal-T5 protein from the host cells. The
CC   methods of the invention can be used for recombinant production of
CC   beta3Gal-T5 for use as a catalyst and for recombinant production of
CC   peptides or proteins with appropriate galactosylation. The beta3Gal-T5
CC   protein can be used to obtain beta1,3-galactosyl glycosylated
CC   saccharides, glycopeptides or glycoproteins.
XX
SQ   Sequence 1011 BP; 247 A; 256 C; 251 G; 257 T; 0 other;

  Query Match             33.5%;  Score 928.6;  DB 21;  Length 1011;
  Best Local Similarity   99.6%;  Pred. No. 1.5e-208;
  Matches  931;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    400 AAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTT 459
          | |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     77 AGATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCTTCTGGTTCTGGGGGCTCTTT 136

Qy    460 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 GTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGA 196

Qy    520 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 AAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCG 256

Qy    580 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    257 TCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGT 316

Qy    640 GGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 699
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    317 GGGGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCA 376

Qy    700 CCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 759
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    377 CCAGCAGTGCAGCGGAAACAAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTA 436

Qy    760 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 TCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAG 496

Qy    820 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 AATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGT 556

Qy    880 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 TCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTT 616

Qy    940 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    617 TCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTG 676

Qy   1000 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 1059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 TCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACG 736

Qy   1060 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 1119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    737 TGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTA 796

Qy   1120 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 1179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    797 AACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGC 856

Qy   1180 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 1239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    857 TCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGA 916

Qy   1240 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 1299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 GGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAG 976

Qy   1300 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGA 1334
          ||||||||||||||||||||||||||||||||||
Db    977 AGAATTCCCGGGGGGAAGATTGTCCGCCTGTCTGA 1011


RESULT 4
AAQ67067
ID   AAQ67067 standard; cDNA to mRNA; 1739 BP.
XX
AC   AAQ67067;
XX
DT   14-MAR-1995  (first entry)
XX
DE   Beta-1,3-galactosyltransferase cDNA.
XX
KW   Beta-1,3-galactosyltransferase; enzyme; saccharide chain; pAMoPRWM1;
KW   KJM-1 cells; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             676..1656
FT                   /*tag= a
FT                   /product= Beta-1,3-galactosyltransferase
XX
PN   JP06181759-A.
XX
```

```
PD   05-JUL-1994.
XX
PF   16-DEC-1992;   92JP-0336436.
XX
PR   16-DEC-1992;   92JP-0336436.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
DR   WPI; 1994-251683/31.
DR   P-PSDB; AAR57433.
XX
PT   Beta-galactosyl-transferase DNA and protein - useful for prodn. of
PT   saccharide chains
XX
PS   Claim 3; Page 22-24; 47pp; Japanese.
XX
CC   This sequence encodes a beta-1,3-galactosyltransferase.  This enzyme
CC   can be used to produce physiologically active saccharide chains and
CC   varients, and for improvement of saccharide chains bound to
CC   physiologically active proteins.  This cDNA represents a fragment
CC   of the plasmid pAMoPRWM1 which was cloned in KJM-1 cells.
XX
SQ   Sequence 1739 BP; 473 A; 407 C; 422 G; 437 T; 0 other;

  Query Match             5.9%;  Score 163.4;  DB 15;  Length 1739;
  Best Local Similarity   55.7%;  Pred. No. 2.2e-28;
  Matches  358;  Conservative    0;  Mismatches  276;  Indels    9;  Gaps    2;

Qy     529 ACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGG 588
Db     866 AATTTCTTATCAACGAGCCCAATAAATGTGAGAAAAACATTCCTTTTCTTGTTATCCTCA 925

Qy     589 TGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAG 648
Db     926 TCAGCACCACTCACAAGGAATTTGATGCCCGTCAGGCAATCAGAGAGACGTGGGGGGATG 985

Qy     649 AGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTG 708
Db     986 AGAACAACTTTAAGGGGATCAAGATAGCCACCCTGTTCCTCCTGGGCAAGAATGCTGATC 1045

Qy     709 CAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGG 768
Db    1046 CTGTTCTCAATCAGATGGTGGAGCAAGAGAGCCAAATCTTCCATGATATCATCGTGGAGG 1105

Qy     769 ATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCC 828
Db    1106 ACTTTATTGACTCCTACCATAACCTTACCCTCAAAACATTAATGGGGATGAGATGGGTGG 1165

Qy     829 ATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATG 888
Db    1166 CCACTTTTTGTTCAAAAGCCAAGTATGTCATGAAAACAGACAGCGACATTTTTGTAAACA 1225

Qy     889 TTGACTATCTGACTGAACTGCTTCTGAA------GAAAAACAGAACAACCAGGTTTTTCA 942
Db    1226 TGGACAATCTTATTTATAAATTACTGAAACCCTCCACCAAGCCACGAAGAAGGTATTTTA 1285

Qy     943 CTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCA 1002
Db    1286 CTGGCTATGTCA---TTAATGGAGGACCGATTCGGGATGTCCGCAGTAAGTGGTATATGC 1342

Qy    1003 GTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGT 1062
Db    1343 CCAGGGATTTGTACCCAGACAGTAACTACCCACCTTTCTGTTCGGGGACTGGCTACATCT 1402

Qy    1063 TTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAAC 1122
Db    1403 TTTCAGCCGATGTAGCTGAACTCATTTACAAGACCTCACTCCACACAAGGCTGCTTCACC 1462

Qy    1123 TGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACAT 1165
Db    1463 TTGAAGACGTATATGTGGGACTGTGTCTTCGAAAGCTGGGCAT 1505

RESULT 5
AAV57905
ID   AAV57905 standard; cDNA; 3518 BP.
XX
AC   AAV57905;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Bovine butyrophilin gene BTF2.
XX
KW   Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW   diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW   BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW   type 1 sodium transport gene; ss.
XX
OS   Bos sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             87..1658
FT                   /*tag= a
FT                   /product= "BTF2"
XX
PN   WO9814466-A1.
XX
PD   09-APR-1998.
XX
PF   30-SEP-1997;   97WO-US17658.
XX
PR   07-MAY-1997;   97US-0852495.
PR   01-OCT-1996;   96US-0724394.
XX
PA   (PROG-) PROGENTIOR INC.
XX
PI   Feder JN,  Kronmal GS,  Lauer PM,  Ruddy DA,  Thomas WJ;
PI   Tsuchihashi Z,  Wolff RK;
XX
DR   WPI; 1998-240014/21.
DR   P-PSDB; AAW78915.
XX
PT   Hereditary haemochromatosis gene products - used to develop products
PT   for the diagnosis and treatment of hereditary disorders in iron
PT   metabolism
XX
PS   Claim 28; Fig 7; 209pp; English.
XX
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   presence of the HFE gene mutation in the genome of the individual. The
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes BTF genes, which are homologues of the milk
CC   protein butyrophilin (BT), and can be used in the production of agonists
CC   and antagonists of BT function. The present sequence encodes BTF2. Also
CC   described are: (1) a RoRet gene which can be used to develop products
CC   for the study, diagnosis and treatment of lupus and Sjogren's syndrome;
CC   and (2) NPT3 and NPT4 genes which are homologues of a type 1 sodium
CC   transport gene, and can similarly be used for hypophosphatemia.
XX
SQ   Sequence 3518 BP; 926 A; 831 C; 868 G; 893 T; 0 other;

  Query Match             4.8%;  Score 133.6;  DB 19;  Length 3518;
  Best Local Similarity   78.9%;  Pred. No. 3e-21;
  Matches  172;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    1;

Qy      12 TTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCAGAGCTGCAGGC 71
Db    3101 TCTCTCCTCTGGAGCTGAGACACTCTTCTTCTTCTGCCCTTGGACATCAGAACTCCTGGC 3160
```

```
Qy      72 TCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGGCCTTTG 131
           |||| |||||||| ||   |||| || ||||| |   |   ||||||||||||||||||
Db    3161 TCTCCGGCCTTTGAACTTCAGGACTTGTACCAGG--AGGCCCTGGGTTCTCAGGCCTTTG 3218

Qy     132 GCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCTTGGACT 191
           || |||||||| |||||||| ||  || |  ||||||||||||| |  | |||  |||
Db    3219 GCTTTGGACTGAGAGTTACACAATCAGCTTCCCTGGTTCTGAGGCTTTCAGACTTAAACT 3278

Qy     192 GAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCA 229
           ||||||  || |  |||||||| |||||||| | ||
Db    3279 GAGCCATGCTACCAGCATCCCAGGGTCTCCAGCCTACA 3316


RESULT 6
AAX87193
ID   AAX87193 standard; cDNA; 1773 BP.
XX
AC   AAX87193;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Human Dendriac cDNA.
XX
KW   Dendriac; Brainiac-2; human; Notch; immune disorder;
KW   neurological disorder; diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..980
FT                   /*tag= a
FT   sig_peptide     20..94
FT                   /*tag= b
FT   mat_peptide     95..977
FT                   /*tag= c
XX
PN   WO9931116-A1.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-US27049.
XX
PR   17-NOV-1998;   98US-0108928.
PR   18-DEC-1997;   97US-0068006.
PR   12-MAR-1998;   98US-0077687.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ebner R,  Endress GA,  Florence KA,  Rosen CA,  Ruben SM;
PI   Soppet DR,  Yu G;
XX
DR   WPI; 1999-430031/36.
DR   P-PSDB; AAY06461.
XX
PT   Human proteins for treating and detecting immune and neurological
PT   disorders
XX
PS   Claim 2; Page 119-120; 132pp; English.
XX
CC   This is the nucleotide sequence of a cDNA clone which codes for
CC   human Dendriac (see AAY06461), also called Brainiac-2, a novel
CC   member of the Brainiac family.  Dendriac cDNA (see also AAX87193) is
CC   deposited as ATCC 203056 and ATCC 209627.  It was initially
CC   discovered in a dendritic cell cDNA library.  Additional clones of
CC   the same gene were identified in NTERA2 cell, adult pulmonary
CC   tissue, salivary gland, ovary, Caco-2 colon adenocarcinoma, smooth
CC   muscle, cerebellum, 8-week-old human mrbyo, haemagiopericytoma,
CC   amygdala, substantia nigra and whole brain cDNA libraries.  Dendriac
CC   nucleic acids are used in the recombinant production of Dendriac
CC   polypeptides, especially the mature protein or epitope-bearing
CC   fragments.  Dendriac and Brainiac proteins are members of the Notch
CC   family of proteins involved in the control of cell division.  The
CC   proteins may be used to detect and/or treat immune system and
CC   neurological disorders.
XX
SQ   Sequence 1773 BP; 568 A; 300 C; 339 G; 566 T; 0 other;

  Query Match             4.7%;  Score 130.4;  DB 20;  Length 1773;
  Best Local Similarity   48.9%;  Pred. No. 1.3e-20;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||   |||||    |||||       |       || || |||||   ||   |  ||
Db     157 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 216

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | ||||||||||||   ||     | |  ||   |   |   ||||||
Db     217 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 276

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | | |      |  |||| |  || |    ||||| ||| |
Db     277 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 336

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                |  || |  |    ||          | ||  | ||   ||
Db     337 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 396

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
             |  | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db     397 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 456

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           | ||||| ||||   |   |||||  |   ||||| ||  | ||   ||  || ||||
Db     457 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 516

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |     |   ||| | ||   ||
Db     517 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 576

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db     577 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 636

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
           | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db     637 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 696

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  | |
Db     697 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 756

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db     757 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 816

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |  |          || |||| |    |  | | | ||
Db     817 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 876

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
           || || |  | ||  |  | | ||  |||   | |     |   |    || |
Db     877 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 936

Qy    1282 ACTACTGGCAGG 1293
             | |||||||
Db     937 CTTTTTGGCAGG 948


RESULT 7
ABK51201
ID   ABK51201 standard; cDNA; 1897 BP.
XX
AC   ABK51201;
XX
DT   30-JUL-2002  (first entry)
XX
```

```
DE   Human cDNA encoding beta1,3-acetylgalactosamine transferase.
XX
KW   Human; ss; gene; beta1,3-acetylgalactosamine transferase;
KW   Gb4 sugar chain.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    109..1101
FT                   /*tag= a
FT                   /note= "This region is specifically claimed in claim 1"
FT   CDS             109..1104
FT                   /*tag= b
FT                   /product= "Beta1,3-acetylgalactosamine transferase"
FT   polyA_signal    1157..1162
FT                   /*tag= c
XX
PN   JP2002085069-A.
XX
PD   26-MAR-2002.
XX
PF   08-SEP-2000; 2000JP-0273835.
XX
PR   08-SEP-2000; 2000JP-0273835.
XX
PA   (SEGK ) SEIKAGAKU KOGYO CO LTD.
PA   (FURU/) FURUKAWA K.
XX
DR   WPI; 2002-378274/41.
DR   P-PSDB; AAU80224.
XX
PT   Preparation of beta-1,3-acetylgalactosamine transferase -
XX
PS   Claim 1; Page 11-13; 15pp; Japanese.
XX
CC   The invention relates to the preparation of beta-1,3-acetylgalactosamine
CC   (a Gb4 sugar chain molecule) comprising introducing a DNA appearing as
CC   ABK51201 (S1) or a DNA hybridising with S1 or a base sequence
CC   complementary to the base sequence or part of these base sequences under
CC   a stringent condition into a cell and growing the cell to express beta-1,
CC   3-acetylgalactosamine transferase and collecting it. Also included is the
CC   preparation of Gb4 sugar chain comprising contacting a polypeptide
CC   comprising the beta-1,3-acetylgalactosamine transferase protein appearing
CC   as AAU80224 (A) or a polypeptide consisting of an amino acid sequence in
CC   which at least one amino acid is replaced, deleted, inserted or
CC   transferred in the amino acid sequence (A) and having enzymatic activity
CC   transferring N-acetylgalactosamine residue from an N-acetylgalactosamine
CC   donor to the C3 site of the galactose residue in Gb3 sugar chain which is
CC   the receptor with an N-acetylgalactosamine donor and Gb3 sugar chain.
CC   The present sequence is the cDNA encoding beta-1,3-acetylgalactosamine
CC   transferase.
XX
SQ   Sequence 1897 BP; 596 A; 327 C; 366 G; 608 T; 0 other;

  Query Match             4.7%;  Score 130.4;  DB 24;  Length 1897;
  Best Local Similarity   48.9%;  Pred. No. 1.3e-20;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||       |      || || |||||   ||   |  ||
Db    281 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 340

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db    341 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 400

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |      |  |||| |  || |    |||||  ||| |
Db    401 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 460

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||               | || |  |    ||         | ||  | ||   ||
Db    461 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 520

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
            |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db    521 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 580

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   |||||  ||  | ||   ||  || ||||
Db    581 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 640

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | ||  |      |    ||| | ||   ||
Db    641 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 700

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    701 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 760

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||       | |   | |    || |||||  |   |   || ||| |||| ||
Db    761 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 820

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    821 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 880

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
            |||   |  ||| ||  | |||||  || | ||| ||| ||||  |  |     | |
Db    881 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 940

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db    941 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1000

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  |  | ||  |||   | |     |   |    || |
Db   1001 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1060

Qy   1282 ACTACTGGCAGG 1293
            |  |||||||
Db   1061 CTTTTTGGCAGG 1072


RESULT 8
AAZ65022
ID   AAZ65022 standard; cDNA; 2095 BP.
XX
AC   AAZ65022;
XX
DT   05-APR-2000  (first entry)
XX
DE   Membrane-bound protein PRO1074 encoding cDNA.
XX
KW   Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW   pharmaceutical; receptor immunoadhesin; gene mapping; ss.
XX
OS   Homo sapiens.
XX
PN   WO9963088-A2.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999;   99WO-US12252.
XX
PR   02-JUN-1998;   98US-0087607.
PR   02-JUN-1998;   98US-0087609.
PR   02-JUN-1998;   98US-0087759.
PR   03-JUN-1998;   98US-0087827.
PR   04-JUN-1998;   98US-0088021.
PR   04-JUN-1998;   98US-0088025.
PR   04-JUN-1998;   98US-0088028.
PR   04-JUN-1998;   98US-0088029.
PR   04-JUN-1998;   98US-0088030.
PR   04-JUN-1998;   98US-0088033.
```

PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
DR WPI; 2000-072883/06.
DR P-PSDB; AAY66685.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
PS Claim 2; Fig 136; 822pp; English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block

```
CC   receptor-ligand interactions. The membrane-bound proteins can also be
CC   employed for screening of potential peptide or small molecule inhibitors
CC   of the relevant receptor/ligand interaction. The PRO encoding sequences
CC   are useful as hybridization probes, in chromosome and gene mapping and in
CC   the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC   will also be useful for the preparation of PRO polypeptides, especially
CC   by recombinant techniques.
XX
SQ   Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

  Query Match             4.7%;  Score 130.4;  DB 21;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 1.4e-20;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
Db    1274 CTTTTTGGCAGG 1285


RESULT 9
AAS46005
```

```
ID   AAS46005 standard; cDNA; 2095 BP.
XX
AC   AAS46005;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human DNA encoding PRO polypeptide sequence #81.
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW   dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168848-A2.
XX
PD   20-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06520.
XX
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   03-MAR-2000; 2000US-187202P.
PR   06-MAR-2000; 2000US-186968P.
PR   14-MAR-2000; 2000US-189320P.
PR   14-MAR-2000; 2000US-189328P.
PR   15-MAR-2000; 2000WO-US06884.
PR   21-MAR-2000; 2000US-190828P.
PR   21-MAR-2000; 2000US-191007P.
PR   21-MAR-2000; 2000US-191048P.
PR   21-MAR-2000; 2000US-191314P.
PR   28-MAR-2000; 2000US-192655P.
PR   29-MAR-2000; 2000US-193032P.
PR   29-MAR-2000; 2000US-193053P.
PR   30-MAR-2000; 2000WO-US08439.
PR   04-APR-2000; 2000US-194449P.
PR   04-APR-2000; 2000US-194647P.
PR   11-APR-2000; 2000US-195975P.
PR   11-APR-2000; 2000US-196000P.
PR   11-APR-2000; 2000US-196187P.
PR   11-APR-2000; 2000US-196690P.
PR   11-APR-2000; 2000US-196820P.
PR   18-APR-2000; 2000US-198121P.
PR   18-APR-2000; 2000US-198585P.
PR   25-APR-2000; 2000US-199397P.
PR   25-APR-2000; 2000US-199550P.
PR   25-APR-2000; 2000US-199654P.
PR   03-MAY-2000; 2000US-201516P.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   05-JUN-2000; 2000US-209832P.
PR   28-JUL-2000; 2000WO-US20710.
PR   22-AUG-2000; 2000US-0644848.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Chen J,  Desnoyers L,  Goddard A,  Godowski PJ,  Gurney AL;
PI   Pan J,  Smith V,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-602746/68.
DR   P-PSDB; AAU29104.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
```

XX
PS Claim 2; Fig 161; 774pp; English.
XX
CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
SQ Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

Query Match 4.7%; Score 130.4; DB 22; Length 2095;
Best Local Similarity 48.9%; Pred. No. 1.4e-20;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||       |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||   ||    | |  ||   |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |     |  |||| |  || |    |||||  ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                  |  || |  |    ||        | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
          |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||    |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||| || |       | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||    | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |      | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
            |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |   |         || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213
```

```
Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  |  | ||   |||   | |      |   |    || |
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
            |  |||||||
Db   1274 CTTTTTGGCAGG 1285
```

RESULT 10
AAF92075
ID AAF92075 standard; cDNA; 2095 BP.
XX
AC AAF92075;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1074 cDNA.
XX
KW Human; PRO protein; mapping; ss.
XX
OS Homo sapiens.
XX
PN WO200116318-A2.
XX
PD 08-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-US23328.
XX
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
PA (GETH) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI; 2001-183260/18.
DR P-PSDB; AAB87543.
XX
PT Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping. -
XX
PS Claim 2; Fig 35; 278pp; English.
XX
CC The present sequence is the coding sequence for a human PRO polypeptide
CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
CC antagonists or anti-PRO antibodies are useful for preparation of a
CC medicament useful in the treatment of a condition which is responsive to
CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
CC protein may also be employed as molecular weight markers for protein
CC electrophoresis. The PRO coding sequence has applications in molecular
CC biology, including use as hybridisation probes, and in chromosome and
CC gene mapping.
XX
SQ Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

Query Match 4.7%; Score 130.4; DB 22; Length 2095;
Best Local Similarity 48.9%; Pred. No. 1.4e-20;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||  |||||   |||||      |      || || |||||  ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | ||||||||||   ||     | |  ||   |   |  ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | | |     |  |||| |  || |    |||| ||| |
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                 |  || |  |    ||         | ||  | ||   ||
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
           |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           | |||||  ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |      |     ||| | ||   ||
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||| || |       | |  || | | | ||   ||
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
           | |  ||      | |  | |    || ||||| |   |   || ||| |||| ||
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
           || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
             |  |||||||
Db    1274 CTTTTTGGCAGG 1285


RESULT 11
AAF44168
ID   AAF44168 standard; cDNA; 2095 BP.
XX
AC   AAF44168;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO1074 (UNQ531) nucleotide sequence SEQ ID NO:208.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000;  2000WO-US00219.
PR   06-JAN-2000;  2000WO-US00376.
PR   11-FEB-2000;  2000WO-US03565.
PR   18-FEB-2000;  2000WO-US04341.
PR   22-FEB-2000;  2000WO-US04414.
PR   24-FEB-2000;  2000WO-US04914.
PR   24-FEB-2000;  2000WO-US05004.
PR   02-MAR-2000;  2000WO-US05841.
PR   15-MAR-2000;  2000WO-US06884.
PR   20-MAR-2000;  2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   P-PSDB; AAB65208.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 2; Fig 136; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

  Query Match             4.7%;  Score 130.4;  DB 22;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 1.4e-20;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||  |||||   |||||      |      || || |||||  ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553
```

```
Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
Db    1274 CTTTTTGGCAGG 1285
```

RESULT 12
AAH15711
ID AAH15711 standard; cDNA; 2168 BP.
XX
AC AAH15711;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14092.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14092; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2168 BP; 634 A; 405 C; 452 G; 677 T; 0 other;

Query Match 4.7%; Score 130.4; DB 22; Length 2168;
Best Local Similarity 48.9%; Pred. No. 1.4e-20;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
Db     571 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 630

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
Db     631 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 690

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
Db     691 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 750

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
Db     751 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 810

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
Db     811 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 870

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
Db     871 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAACGCCAAGTACGTAATGA 930
```

```
Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db     931 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 990

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||| || |       | |  || | | | ||   ||
Db     991 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 1050

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
             | |   ||       | |   | |    || ||||| |   |   || ||| |||| ||
Db    1051 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTCCCCTCCATACT 1110

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1111 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1170

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    1171 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1230

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |   |          || |||| |    |  | | ||
Db    1231 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1290

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
             || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1291 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1350

Qy    1282 ACTACTGGCAGG 1293
             |  |||||||
Db    1351 CTTTTTGGCAGG 1362


RESULT 13
AAX87192
ID   AAX87192 standard; cDNA; 2189 BP.
XX
AC   AAX87192;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Human Dendriac cDNA.
XX
KW   Dendriac; Brainiac-2; human; Notch; immune disorder;
KW   neurological disorder; diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             435..1394
FT                   /*tag=  a
FT   sig_peptide     435..509
FT                   /*tag=  b
FT   mat_peptide     510..1391
FT                   /*tag=  c
XX
PN   WO9931116-A1.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-US27049.
XX
PR   17-NOV-1998;   98US-0108928.
PR   18-DEC-1997;   97US-0068006.
PR   12-MAR-1998;   98US-0077687.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ebner R,  Endress GA,  Florence KA,  Rosen CA,  Ruben SM;
PI   Soppet DR,  Yu G;
XX
DR   WPI; 1999-430031/36.
DR   P-PSDB; AAY06461.
XX
PT   Human proteins for treating and detecting immune and neurological
PT   disorders
XX
PS   Claim 2; Page 110-112; 132pp; English.
XX
CC   This is the nucleotide sequence of cDNA clone HFVIF40, which codes
CC   for human Dendriac (see AAY06461), also called Brainiac-2, a novel
CC   member of the Brainiac family.  Dendriac cDNA is deposited as
CC   ATCC 203056 and ATCC 209627.  The cDNA was initially discovered in
CC   a cDNA library derived from dendritic cells.  Additional clones of
CC   the same gene were identified in NTERA2 cell, adult pulmonary
CC   tissue, salivary gland, ovary, Caco-2 colon adenocarcinoma, smooth
CC   muscle, cerebellum, 8-week-old human mrbyo, haemagiopericytoma,
CC   amygdala, substantia nigra and whole brain cDNA libraries.  Dendriac
CC   nucleic acids are used in the recombinant production of Dendriac
CC   polypeptides, especially the mature protein or epitope-bearing
CC   fragments.  Dendriac and Brainiac proteins are members of the Notch
CC   family of proteins involved in the control of cell division.  The
CC   proteins may be used to detect and/or treat immune system and
CC   neurological disorders.
XX
SQ   Sequence 2189 BP; 657 A; 402 C; 453 G; 677 T; 0 other;

  Query Match             4.7%;  Score 130.4;  DB 20;  Length 2189;
  Best Local Similarity   48.9%;  Pred. No. 1.4e-20;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||   |||||    |||||       |       || || |||||   ||   |  ||
Db     571 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 630

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | |||||||||||   ||    | |  ||   |   |  ||||||
Db     631 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 690

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | | |      |  |||| |  || |    ||||| ||| |
Db     691 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 750

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                |  || | |   ||         | ||  | ||   ||
Db     751 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 810

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
             |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db     811 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 870

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db     871 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 930

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db     931 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 990

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||| || |       | |  || | | | ||   ||
Db     991 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 1050

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
             | |   ||       | |   | |    || ||||| |   |   || ||| |||| ||
Db    1051 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1110

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1111 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1170

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    1171 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1230
```

```
Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |   |         || |||| |    |  | |  ||
Db   1231 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1290

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  |  | ||  |||  | |     |   |    || |
Db   1291 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1350

Qy   1282 ACTACTGGCAGG 1293
          |  |||||||
Db   1351 CTTTTTGGCAGG 1362
```

RESULT 14
AAX51967
ID AAX51967 standard; DNA; 333 BP.
XX
AC AAX51967;
XX
DT 22-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO: 181.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906552-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01236.
XX
PR 01-AUG-1997; 97US-0905223.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153782/13.
DR P-PSDB; AAY13167.
XX
PT New isolated brain-derived nucleic acids - used to develop products
PT which may have cytokine, immune, regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Claim 1; Page 353; 577pp; English.
XX
CC AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12987 to
CC AAY13219, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 333 BP; 66 A; 95 C; 80 G; 92 T; 0 other;

```
  Query Match             4.5%;  Score 124.8;  DB 20;  Length 333;
  Best Local Similarity   74.2%;  Pred. No. 1.5e-19;
  Matches  184;  Conservative    0;  Mismatches   62;  Indels    2;  Gaps    2;

Qy      24 AGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCAGAGCTGCAGGCTCTCTGGCCTTT 83
           |||| ||| |||||||| |||||| || ||||||||||| |  |||||||||  ||||||
Db       2 AGCTAGGACATTCTTCTCCTCCTGGCCCTGGACATCAGAACCCCAGGCTCTCCAGCCTTT 61

Qy      84 GGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGGCCTTTGGCCTTGGACTGA 143
           ||||   |||| | | | ||||||   | |||||||||| |||||||||| |||| |||||
Db      62 GGACTTCAGGACTGACA-CAAGCAACCTGCTGGGTTCTTAGGCCTTTGG-CTTGTACTGA 119

Qy     144 TAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCTTGGACTGAGCCACACTCC 203
           | |||||||||  || |  ||||| ||||| || |||| ||||||  |||||||| || |
Db     120 GACTTACACCATCAGCTTCCCTGGTCCTGAGACTTTTGGACTTGGATTGAGCCACGCTAC 179

Qy     204 TGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCACGTGATTCCTGTCAGAATCACCAT 263
           ||||||||||  ||||||||||||||        |  ||||   | |  |||  ||   |
Db     180 TGGCATCCCAGGATCTCCAGCTTGCAGACAGCCTGTCGTGGGACTTCACAGCCTCCATAA 239

Qy     264 TTTTGGTA 271
           || | | |
Db     240 TTATAGAA 247
```

RESULT 15
AAX35711
ID AAX35711 standard; cDNA; 2420 BP.
XX
AC AAX35711;
XX
DT 09-JUL-1999 (first entry)
XX
DE cDNA encoding a protein identified by the signal sequence trap method.
XX
KW Signal sequence trap method; SST method; immunisation; inhibition;
KW infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW activin activity; inhibin activity; chemokine activity;
KW cytokine activity; blood coagulation regulation; agonist; antagonist;
KW metabolic disorder; hormonal disorder; immune disorder;
KW severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
KW wound; ss.
XX
OS Homo sapiens.
XX
PN WO9918126-A1.
XX
PD 15-APR-1999.
XX
PF 06-OCT-1998; 98WO-JP04514.
XX
PR 07-OCT-1997; 97JP-0274674.
XX
PA (ONOY) ONO PHARM CO LTD.
XX
PI Fukushima D, Shibayama S, Tada H;
XX
DR WPI; 1999-277254/23.
DR P-PSDB; AAY02366.
XX
PT Polypeptides identified by the signal sequence trap method from a
PT human cDNA library
XX
PS Claim 5; Page 110-113; 281pp; Japanese.
XX
CC AAX35694-X35747 represent cDNA sequences that encode novel polypeptides
CC (AAY02358-84) which are identified from a human placental cDNA library
CC by the signal sequence trap (SST) method. The polypeptides have a
CC broad range of physiological activity, including immunisation against
CC and inhibition of infections, allergies and cancer; regulation of tissue
CC formation and repair; activin/inhibin activity; chemokine/cytokine
CC activity; blood coagulation regulation; and receptor/ligand agonist
CC or antagonist activity. The polypeptides can be used for prevention

```
CC   and treatment of disorders including infections by bacteria, yeasts and
CC   viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC   immune disorders (including severe combined immunodeficiency (SCID)
CC   and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX
SQ   Sequence 2420 BP; 853 A; 458 C; 451 G; 658 T; 0 other;

  Query Match             4.5%;  Score 124;  DB 20;  Length 2420;
  Best Local Similarity   49.0%;  Pred. No. 4.7e-19;
  Matches  424;  Conservative    0;  Mismatches  430;  Indels   12;  Gaps    3;

Qy     551 AGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTT 610
           | | |||    ||| |  ||| ||  |  | || ||  |  |  ||   |    ||| |
Db    1129 AAAATGCCAAGAGAAAAGTCCTTTTTTAATACTACTAATAGCTGCAGAGCCTGGACAAAT 1188

Qy     611 GGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCA 670
           |  |   | |  || || ||||| || ||||| || || ||  | |     || |  ||
Db    1189 AGAAGCTAGAAGAGCTATTCGGCAAACTTGGGGCAATGAAAGTCTAGCACCTGGTATTCA 1248

Qy     671 GCTGAAGACATTCTTCCTCCTGGGGACCACCA-GCAGTGCAGCGGAAACGAAAGAGGTGG 729
           | |  | | | ||  |  ||||    |  |   ||   |  ||  ||     | |||
Db    1249 AATCACAAGAATATTTTTGTTGGGCTTAAGTATTAAGCTAAATGGCTACCTTCAACGTGC 1308

Qy     730 A-----CCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTA 784
           |       | || |||   | | |    || || || ||  |||| | | ||||    ||
Db    1309 AATACTGGAAGAAAGCAGACAATATCATGATATAATTCAACAGGAATACTTAGATACGTA 1368

Qy     785 TTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCA 844
           || ||| ||||| | || ||  | |||||||  | |||||       |  ||||| ||
Db    1369 CTATAATTTGACCATTAAAACACTAATGGGCATGAACTGGGTTGCAACATACTGTCCACA 1428

Qy     845 GGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGA 904
           | | ||| |||||||| |||   ||||||||  ||||   ||| ||| | |   |
Db    1429 TATTCCATATGTTATGAAAACTGACAGTGACATGTTTGTCAACACTGAATATTTAATCAA 1488

Qy     905 ACTGCTTCTGAAGAAAAACAGAACAACCAG------GTTTTTCACTGGCTTCTTGAAACT 958
           | | ||||||  | |      |  ||||       | |||||||||| | | | |  |
Db    1489 TAAGTTACTGAAGCCAGATCTGCCTCCCAGACATAACTATTTCACTGGTTACCTAATGCG 1548

Qy     959 CAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCC 1018
           |    ||||   | |   |  ||||||||||   |      |      || ||
Db    1549 AGGATATGCACCCAATCGAAACAAAGATAGCAAGTGGTACATGCCACCAGACCTCTACCC 1608

Qy    1019 GTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGC 1078
           | ||  | || ||    ||||| || || || || || || |||||||| ||  ||||
Db    1609 AAGTGAGCGTTATCCTGTCTTCTGTTCTGGAACTGGTTATGTTTTTTCTGGAGATCTGGC 1668

Qy    1079 GAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGT 1138
           || | |  || || ||      |  |||      |  |  ||||||| || | |||
Db    1669 AGAAAAGATTTTTAAAGTTTCTTTAGGTATCCGCCGTTTGCACTTGGAAGATGTATATGT 1728

Qy    1139 GGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTT 1198
           ||| |||| || |  | | |||  ||       |      |  | |    ||      ||
Db    1729 AGGGATCTGTCTTGCCAAGTTGAGAATTGATCCTGTACCCCCTCCCAATGAGTTTGTGTT 1788

Qy    1199 TTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTT 1258
           ||   || |  |    |  ||    ||    | |||     |  |  |||  ||
Db    1789 CAATCACTGGCGAGTCTCTTATTCGAGCTGTAAATACAGCCACCTAATTACCTCTCATCA 1848

Qy    1259 CATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATTCCCGGGGGGAAGA 1318
           || ||||| |    ||  |  | ||||||  |   | || |  |       |    | |
Db    1849 GTTCCAGCCTAGTGAACTGATAAAATACTGGAACCATTTACAACAAAATAAGCACAATGC 1908

Qy    1319 TTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCACATCCGGACAAGTTTCAGATAACCCGTG 1378
           ||| |    |    |  ||      |||||  |   | ||  |  |    |   |
Db    1909 CTGTGCCAACGCAGCAAAAGAAAAGGCAGGCAGGTATCGCCACCGTAAACTACATTAGAA 1968

Qy    1379 GGGATAGTTTTTGCTAGATTTTGGAA 1404
           || | |||||  |  | | || ||
Db    1969 AAGACAATTTTTTTTCAAATGTGCAA 1994

Search completed: April 11, 2003, 20:07:01
```

Job time : 520.132 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 15:35:54 ; Search time 3089.38 Seconds
(without alignments)
14547.424 Million cell updates/sec

Title: US-09-914-152-2
Perfect score: 2775
Sequence: 1 gtgaattcctctttctctgc.....aaaaaaaaaaaaaaaaaaa 2775

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	505.4	18.2	507	9	AI739518	AI739518 wi23a06.x
c	2	455	16.4	455	10	AW136254	AW136254 UI-H-BI1-
c	3	374.4	13.5	434	12	BF748181	BF748181 RC3-BN041
c	4	326	11.7	338	9	AJ003597	AJ003597 AJ003597
	5	238.4	8.6	300	9	AU100432	AU100432 AU100432
c	6	234	8.4	400	17	AZ768940	AZ768940 1M0569K09
	7	234	8.4	528	17	AQ266346	AQ266346 RPCI11-72
c	8	184	6.6	1101	17	CNS05BT1	AL330238 Tetraodon
	9	179	6.5	636	13	BG965086	BG965086 602829157
	10	135	4.9	1514	11	AK003837	AK003837 Mus muscu
c	11	133.6	4.8	491	14	BQ021377	BQ021377 UI-H-DH1-
	12	133.6	4.8	3199	11	BC013258	BC013258 Homo sapi
	13	132.2	4.8	596	13	BJ501651	BJ501651 BJ501651
	14	128.8	4.6	3564	11	BC028571	BC028571 Homo sapi
	15	127	4.6	233	10	BB595953	BB595953 BB595953
	16	123.4	4.4	444	13	BM403751	BM403751 zam6155 Z
c	17	122.8	4.4	400	9	AI311905	AI311905 qo89e07.x
c	18	122.8	4.4	462	9	AI742535	AI742535 wg55b03.x
c	19	122.8	4.4	525	9	AI820087	AI820087 wg56b08.x
	20	122.6	4.4	879	14	BQ895238	BQ895238 AGENCOURT
c	21	122.4	4.4	329	9	AI183500	AI183500 qd42e07.x
c	22	121	4.4	651	14	BQ180890	BQ180890 UI-H-EU0-
	23	120.8	4.4	399	17	AQ475568	AQ475568 CITBI-E1-
	24	120.4	4.3	701	12	BF131199	BF131199 601819426
	25	120	4.3	688	9	AF063529	AF063529 AF063529
	26	120	4.3	688	9	AI110655	AI110655 HA0085 Hu
	27	117.2	4.2	965	17	CNS05IB2	AL338663 Tetraodon
c	28	116.8	4.2	558	14	BQ363849	BQ363849 RC3-ST018
c	29	116.2	4.2	690	17	AG130187	AG130187 Pan trogl
	30	116	4.2	343	12	BF759672	BF759672 CM1-CT060
c	31	115.4	4.2	1021	17	CNS044T4	AL274513 Tetraodon
c	32	114.8	4.1	574	10	BE652985	BE652985 UI-M-AM1-
c	33	113.4	4.1	664	17	AQ346244	AQ346244 RPCI11-10
	34	113.2	4.1	220	14	F05486	F05486 HSC0CF041 n
	35	112.6	4.1	475	17	AQ195633	AQ195633 RPCI11-60
	36	112.6	4.1	491	17	AQ282423	AQ282423 RPCI11-91
	37	112.4	4.1	247	14	T08840	T08840 EST06732 In
c	38	112.4	4.1	632	17	AQ277749	AQ277749 CITBI-E1-
	39	112	4.0	664	14	BQ574062	BQ574062 UI-H-EZ0-
	40	112	4.0	725	14	BQ448351	BQ448351 UI-H-EU1-
	41	111.6	4.0	730	17	AG029969	AG029969 Pan trogl
	42	110.2	4.0	380	17	AQ418581	AQ418581 RPCI-11-2
	43	110.2	4.0	417	17	AQ415767	AQ415767 RPCI-11-2
	44	110.2	4.0	586	17	AQ372907	AQ372907 RPCI11-15
c	45	109.6	3.9	574	10	BE272364	BE272364 601125944

ALIGNMENTS

RESULT 1
AI739518/c
LOCUS AI739518 507 bp mRNA linear EST 20-DEC-1999
DEFINITION wi23a06.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2391058 3', mRNA sequence.
ACCESSION AI739518
VERSION AI739518.1 GI:5101499
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1331 Std Error: 0.00
Seq primer: -40UP from Gibco

```
          High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2391058"
                     /clone_lib="NCI_CGAP_Co16"
                     /tissue_type="colon tumor, RER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Co10 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1057416-1061255, and 1144584-1145351).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      154 a    122 c    115 g    116 t
ORIGIN

  Query Match             18.2%;  Score 505.4;  DB 9;  Length 507;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-77;
  Matches  506;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    2223 CAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACTCAGATGGGGTTCA 2282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     507 CAGGAGCACTTTAGGGCTCTCAGTTCAAACTGAAGGACAGTTGAACTCAGATGGGGTTCA 448

Qy    2283 TGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAAGG 2342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 TGTGGGATTCTGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAAGG 388

Qy    2343 ACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCCAC 2402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     387 ACCCCTCGGGCTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCCAC 328

Qy    2403 TTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAACTG 2462
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 TTCCCAAGCCTGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAACTG 268

Qy    2463 AACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTTTG 2522
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 AACTTTGTGGTTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTTTG 208

Qy    2523 TGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATTCA 2582
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 TGCGAGCCTGTGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATTCA 148

Qy    2583 GCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCTTT 2642
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147 GCCATGGTCACGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCTTT 88

Qy    2643 TAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGGCTGTG 2702
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      87 TAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAACAGGCCTCACATGGCTGTG 28

Qy    2703 TCACATATAAATGTTGGACTAAACTCT 2729
           |||||||||||||||||||||||||||
Db      27 TCACATATAAATGTTGGACTAAACTCT 1


RESULT 2
AW136254/c
LOCUS       AW136254                455 bp    mRNA    linear   EST 29-OCT-1999
DEFINITION  UI-H-BI1-act-e-12-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
            IMAGE:2715383 3', mRNA sequence.
ACCESSION   AW136254
VERSION     AW136254.1  GI:6140387
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2715383"
                     /clone_lib="NCI_CGAP_Sub3"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NCI_CGAP_Sub3 library is a subtracted library derived from
                     the  NCI_CGAP_Sub1 library, which is a subtracted library
                     derived from BI.  BI constitutes a mixture of 21
                     normalized or subtracted NCI_CGAP  libraries: NCI_CGAP_Co4
                     , NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
                     NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
                     NCI_CGAP_Kid3,  NCI_CGAP_Kid11, NCI_CGAP_Lym2,
                     NCI_CGAP_Br2, NCI_CGAP_Co8,  NCI_CGAP_CLL1, NCI_CGAP_Lei12,
                     NCI_CGAP_Brn23, NCI_CGAP_Lu5,  NCI_CGAP_Lu24,
                     NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
                     NCI_CGAP_Brn25.  These 21 libraries were pooled and a
                     single-stranded DNA  preparation of the resulting mixture
                     was used as a tracer in a subtractive  hybridization with
                     a driver whose composition is detailed below:
                     NCI_CGAP_Kid3 pool 1    LLAM 3334-3337, 3682-3683,
                     3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775
                     , 1500552-1502855); NCI_CGAP_Kid5 pool 1    LLAM 3338-3342
                     , 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831,
                     1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1
                     LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,
                     1520904-1522439); NCI_CGAP_GC4 pool 1    LLAM 3164-3167,
                     3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,
                     1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
                     LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs
                     985608-986759, 1101192-1101959, 1217928-1220615);
                     NCI_CGAP_Co10 pool 1    LLAM 2644-2653, 2871-2872 (IMAGE
                     CloneIDs 1057416-1061255, 1144584-1145351). Subtraction
                     was performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To  Facilitate Gene Discovery. Genome Research
                     6, 791-806.
                     TAG_LIB=NCI_CGAP_Co10
                     TAG_TISSUE=colon
                     TAG_SEQ=AAACG"
BASE COUNT      137 a    102 c    101 g    115 t
ORIGIN

  Query Match             16.4%;  Score 455;  DB 10;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-68;
  Matches  455;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2293 TGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAAGGACCCCTCGGG 2352
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     455 TGGGAGCTTTCTGGGAATTCAGTTGGAGTCAAGTCAGGATGCTCTCAAGGACCCCTCGGG 396

Qy    2353 CTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCC 2412
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     395 CTCAGAGCCCTAAAGTGGGCCCTGGTGAAGCAGGGTGGTCCTGCGTCCACTTCCCAAGCC 336

Qy    2413 TGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAACTGAACTTTGTGG 2472
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 TGAGCCAAGCTCATCTTCATTGAATGTCTCATTTGGCCGAGGAACAACTGAACTTTGTGG 276

Qy    2473 TTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTG 2532
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 TTTGCTGTTTAGCCTTCAGTTTGCTCCGCTGCCTCCTACCCAGAGGTTTGTGCGAGCCTG 216

Qy    2533 TGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATTCAGCCATGGTCA 2592
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 TGTTGCAGGGTTGTATAAAACCAAGGTACTTCGTTAGTTTTGCCCATTCAGCCATGGTCA 156

Qy    2593 CGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCTTTTAATTTTTTA 2652
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155 CGTGACATGCAAAGTAATCTTGCTCCTAATTATAGAAATGATTTTTCTTTTAATTTTTTA 96

Qy    2653 CTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGGCTGTGTCACATATAA 2712
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95 CTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGGCTGTGTCACATATAA 36

Qy    2713 ATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAA 2747
           |||||||||||||||||||||||||||||||||||
Db      35 ATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAA 1
```

```
RESULT 3
BF748181/c
LOCUS       BF748181                434 bp    mRNA    linear   EST 10-JAN-2001
DEFINITION  RC3-BN0411-021000-021-c04 BN0411 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF748181
VERSION     BF748181.1  GI:12074857
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-BN0411-
            021000-021-c04&t3=2000-10-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 24
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BN0411"
                     /dev_stage="Adult"
                     /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      104 a    117 c    130 g     83 t
ORIGIN
```

```
  Query Match             13.5%;  Score 374.4;  DB 12;  Length 434;
  Best Local Similarity   97.2%;  Pred. No. 9e-55;
  Matches  381;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy     959 CAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCC 1018
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     434 CAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCC 375

Qy    1019 GTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGC 1078
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     374 GTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGC 315

Qy    1079 GAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGT 1138
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     314 GAGTCAGGTGTACCATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGT 255

Qy    1139 GGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTT 1198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     254 GGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTT 195

Qy    1199 TTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTT 1258
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     194 TTTTCCCGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTT 135

Qy    1259 CATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAGAATTCCCGGGGGGAAGA 1318
           |||||||||||||||| |||||||||||||| ||||||||||||||||||||||||||||
Db     134 CATCAAGCCTCGGACTCCCTTGGACTACTGGCCGGCTCTAGAGAATTCCCGGGGGGAAGA 75

Qy    1319 TTGTCCGCCTGTCTGAGGGGAGCCCAGAGGCA 1350
           |||||||||||||||||||   || ||   ||
Db      74 TTGTCCGCCTGTCTGAGGGAGCCCGAGGCACA 43
```

```
RESULT 4
AJ003597/c
LOCUS       AJ003597                338 bp    mRNA    linear   EST 04-DEC-1997
DEFINITION  AJ003597 Selected chromosome 21 cDNA library Homo sapiens cDNA
            clone MPIpl9-12J9, mRNA sequence.
ACCESSION   AJ003597
VERSION     AJ003597.1  GI:2578270
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Szulzewsky,I., Hunt,E., Nguyen,M., Korn,B., Roehrdanz,B., Lehrach
            ,H. and Yaspo,M.L.H.
  TITLE     An integrated transcript map for the whole human chromosome 21
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Yaspo, M.-L.
            Max Planck Institut fuer Molekulare Genetik
            Ihnestrasse 73, D14195 Berlin-Dahlem, Germany.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="21q"
                     /clone="MPIpl9-12J9"
                     /clone_lib="Selected chromosome 21 cDNA library"
                     /note="Swaroop et al. (1991) Nucleic Acids Res. 19: 1954."
BASE COUNT       76 a     92 c     82 g     87 t      1 others
ORIGIN
```

```
  Query Match             11.7%;  Score 326;  DB 9;  Length 338;
  Best Local Similarity   98.2%;  Pred. No. 2.1e-46;
```

```
  Matches  329;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    416 GATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGTTTGTATTTTAGCAT 475
          || |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    335 GACGAGATTGATGTATATTTGCNTTCTGGTTCTGGGGGCTCTTTGTTTGTATTTTAGCAT 276

Qy    476 GTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAAGACGGGAACTTCCT 535
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    275 GTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAAGACGTGAACTTCCT 216

Qy    536 TAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTC 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215 TAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTC 156

Qy    596 ATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGAT 655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 ATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGAC 96

Qy    656 GGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGA 715
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     95 GGTGAAGGGAAGGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGA 36

Qy    716 AACGAAAGAGGTGGACCAGGAGAGCCAGCGACACG 750
          ||| ||||||||||||||||||||||||||||||
Db     35 AACAAAAGAGGTGGACCAGGAGAGCCAGCGACACG 1


RESULT 5
AU100432
LOCUS       AU100432                 300 bp    mRNA    linear   EST 05-APR-2001
DEFINITION  AU100432 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            colc0014 similar to Homo sapiens mRNA for UDP-Gal:GlcNAc
            beta13-galactosyltransferase 5, mRNA sequence.
ACCESSION   AU100432
VERSION     AU100432.1  GI:13551561
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,K., Suyama,A. and Sugano,S.
  TITLE     In silico mapping of the 5'-ends of human mRNAs using full-length
            enriched and 5'-end enriched cDNA libraries constructed by
            Oligo-capping method
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="colc0014"
                     /clone_lib="Sugano Homo sapiens cDNA library"
                     /note="Differential display comparison of untreated and
                     dimethylfumarate treated U937 cells"
BASE COUNT       57 a     75 c     68 g     85 t     15 others
ORIGIN

  Query Match             8.6%;  Score 238.4;  DB 9;  Length 300;
  Best Local Similarity   92.7%;  Pred. No. 2.7e-31;
  Matches  242;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      1 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCA 60
          |||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||||

Db     40 GTGAATTCCTCTTTCTCTGCTGGAGCTGGGATATACTTCTNCTCCTGCCCTTGGACATCA 99

Qy     61 GAGCTGCAGGCTCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTC 120
          |||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    100 TAGCTGCTGGCTCTCTGGNCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTC 159

Qy    121 TCAGGCCTTTGGCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTT 180
          ||||||||| |||||||||||||||| |||| ||||| |||||||||||||| || |||
Db    160 TCAGGCCTTNGGCCTTGGACTGATAGTNACACNATTGGNATATCTGGTTCTGATGCNCTT 219

Qy    181 GGTCTTGGACTGAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCAC 240
          |||||||||   | |||||||| | |||||||| |||||||||||||| ||||||||||
Db    220 GGTCTTGGANNNATCCACACTCNTNGCATCCCANCGTCTCCAGCTTGCTTGGCCTGTCAC 279

Qy    241 GTGATTCCTGTCAGAATCACC 261
          |||||||||||||||||||||
Db    280 GTGATTCCTGTCAGAATCACC 300


RESULT 6
AZ768940/c
LOCUS       AZ768940                 400 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  1M0569K09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0569K09 F, DNA sequence.
ACCESSION   AZ768940
VERSION     AZ768940.1  GI:12888559
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0569  row: K  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 400.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0569K09"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
```

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 88 a 107 c 104 g 101 t
ORIGIN

Query Match 8.4%; Score 234; DB 17; Length 400;
Best Local Similarity 75.8%; Pred. No. 1.4e-30;
Matches 303; Conservative 0; Mismatches 95; Indels 2; Gaps 1;

```
Qy    491 TTTCAAAGAACAGTCCTTTGTTTACAAGAAAG--ACGGGAACTTCCTTAAGCTCCCAGAT 548
          ||||| |||||   | |||||||  |  ||||  ||||||| |||||| || | || |||
Db    400 TTTCAGAGAACTCCCGTTTGTTTTAAGAAAAGTCACGGGAAGTTCCTTCAGATTCCGGAT 341

Qy    549 ACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAG 608
          | |||||||| |||||  || || ||||| || ||||||||||| ||||| ||||| |||
Db    340 ATAGACTGCAAGCAGAAGCCGCCTTTCCTGGTGCTGCTGGTGACGTCATCTCACAAGCAG 281

Qy    609 TTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAG 668
          ||||  |  |||||||||||||  |||||||||| | |||||     ||| ||||  |
Db    280 CTGGCCGCTCGCATGGCCATCCGCAAGACGTGGGGTAGAGAGACATCTGTGCAGGGCCAA 221

Qy    669 CAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGTG 728
          ||| ||| ||| |||||||| |||||||| ||  |||  | | ||| | | | |
Db    220 CAGGTGAGGACCTTCTTCCTTCTGGGGACCTCCGACAGCACCGAGGAGATGGACGCCACA 161

Qy    729 GACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTAC 788
            || ||||||| |||  ||| | ||||||||||||||||| |||   || | ||| | |
Db    160 ACCCTGGAGAGCGAGCAGCACCGCGACATTATCCAGAAGGACTTCAAGGATGCCTACTTC 101

Qy    789 AATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCG 848
          || |||||||||| |||||||||| || ||||||||| | | ||||||||||||| |
Db    100 AACCTGACCCTGAAAACCATGATGGGTATGGAATGGGTCTACCACTTTTGTCCTCAGACA 41

Qy    849 GCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATG 888
          || |  |||||||||| ||||| ||||||||  | ||||
Db     40 GCTTACGTGATGAAAACGGACTCTGACATGTTTGTGAATG 1
```

RESULT 7
AQ266346
LOCUS AQ266346 528 bp DNA linear GSS 27-APR-1999
DEFINITION RPCI11-72B14.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-72B14, DNA sequence.
ACCESSION AQ266346
VERSION AQ266346.1 GI:3793950
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 528)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .528
/organism="Homo sapiens"
/db_xref="GDB:7527301"
/db_xref="taxon:9606"
/clone="RPCI-11-72B14"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 101 a 146 c 104 g 175 t 2 others
ORIGIN

Query Match 8.4%; Score 234; DB 17; Length 528;
Best Local Similarity 97.5%; Pred. No. 1.2e-30;
Matches 237; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy      6 TTCCTCTTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCAGAGCT 65
          ||||||| ||||||||||||||||||||||||||||||||||||||||||  |||||
Db      1 TTCCTCTNTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATTTGAGCT 60

Qy     66 GCAGGCTCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGG 125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCAGGCTCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGG 120

Qy    126 CCTTTGGCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCT 185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCTTTGGCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCT 180

Qy    186 TGGACTGAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCACGTGAT 245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db    181 TGGACTGAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCACGGTAT 240

Qy    246 TCC 248
          | |
Db    241 TTC 243
```

RESULT 8
CNS05BT1/c
LOCUS CNS05BT1 1101 bp DNA linear GSS 26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 013P23 of library C from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL330238
VERSION AL330238.1 GI:8223860
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1101)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission

```
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="013P23"
                     /clone_lib="C"
                     /note="Genoscope sequence ID : C0CC013CH12T1-end : T3"
BASE COUNT      235 a    239 c    325 g    283 t     19 others
ORIGIN

  Query Match             6.6%;  Score 184;  DB 17;  Length 1101;
  Best Local Similarity   58.5%;  Pred. No. 3.4e-22;
  Matches  354;  Conservative    0;  Mismatches  242;  Indels    9;  Gaps    2;

Qy     568 CTCCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCA 627
           |||| ||||| ||| | ||  | | | |  | |||||  |||| |  | |||   |||||
Db     663 CTCCTTTCCTGGTCATCCTCATCAGCACCACGCACAAGGAGTTTGACGCGCGGCAGGCCA 604

Qy     628 TCCGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCC 687
           ||||| |||| |||||| | ||||| |   | | |      |   |     ||  |||| |
Db     603 TCCGGGAGACCTGGGGGGACGAGAGCACCTTCACCGACGTCCGCATTCTCACCGTCTTTC 544

Qy     688 TCCTGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGAC 747
           | ||||| |  | || |  || ||      |   |   ||||| ||||||||||||
Db     543 TNCTGGGCAGGAACACCGATGAAGTCCTCAACCAGATGGTGGAGCAGGAGAGCCAGATCT 484

Qy     748 ACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCA 807
             |  |||||  | . | || | ||| | ||   |||  |||| || || || || ||
Db     483 TCCATGACATCGTGATGGAGAACTTCATCGATTCCTACCACAACCTCACGCTCAAAACTT 424

Qy     808 TGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAG 867
           ||||||||||    |||||      ||| || ||  | ||    | |||||||| ||||
Db     423 TGATGGGCATGCGCTGGGTGGCCACCTTCTGCCCCAAAGCTCAATATGTGATGAAGACAG 364

Qy     868 ACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAG------AAAA 921
           ||   ||||| ||| ||||  |  || |||| |    |   ||| ||||||      |  |
Db     363 ACAGCGACATCTTCGTCAACATGNACAATCTNATCTACAAGCTCCTGAAGCCCAGCACCA 304

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           |    |  |  |||| |||||| |||| | | |   | || |     || ||  || |
Db     303 AGCCGAGGAGGAGGTATTTCACGGGCTACGTTA---TAAACGGCGGACCGATACGGGACA 247

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
           |||  |||||||    |   |      |||||   |  ||  ||||| || ||||
Db     246 TGCGCAGTAAGTGGTACATGTCGAGGGACTTGTATCCCGAGAGCAAATACCCGCCGTTCT 187

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||||  ||||| |||||||||| || | |||||||||     |   | || ||   ||||
Db     186 GCTCANGCACCNGCTACGTGTTCTCAGCCGACGTGGCCGANCTCATCTATAAGACCTCCC 127

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |  |      |    |  | ||||| |||||||   |||| || ||  |  | | ||||
Db     126 TGCACACAAGACTGTTGCACCTGGAGGACGTGTACNTGGGACTGTGTTTGCATAAGCTGG 67

Qy    1162 ACATC 1166
           ||||
Db      66 GCATC 62


RESULT 9
BG965086
LOCUS       BG965086               636 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602829157F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4983830 5',
            mRNA sequence.
ACCESSION   BG965086
VERSION     BG965086.1  GI:14352723
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10989  row: a  column: 15
            High quality sequence start: 17
            High quality sequence stop: 634.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4983830"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      155 a    172 c    155 g    154 t
ORIGIN

  Query Match             6.5%;  Score 179;  DB 13;  Length 636;
  Best Local Similarity   66.2%;  Pred. No. 3.2e-21;
  Matches  321;  Conservative    0;  Mismatches  155;  Indels    9;  Gaps    4;

Qy     335 ATTTCCTCTTGGCATTTACACTGTGGCTTTAGCTTTCAAACCAGAGGTTCCTCTTACCCA 394
           |||||||||||||| | ||  |||||| || | | |    || |  |    ||||||| |
Db     159 ATTTCCTCTTGGCTTCTATCCTGTGGTTTCATCCTCTGGACGAACGTGAGCTCTTACTCG 218

Qy     395 GCAAAAAATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGC 454
           || ||| ||||||  | |||||  ||  |  | ||    | | |||  |  |||| ||
Db     219 GC-AAAGATGGCTCACATGAAGACAAGGCTAGTTTACGCCTCCATTCTCATGATGGGCGC 277

Qy     455 TCTTTGTTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTA 514
             || || ||||| || |||||| |   | | |         ||   | ||  ||| |
Db     278 ACTCTGCTTGTACTTCAGCATGGATTCTTTCA------GAGAACTCCCGTTGTTTTTAAC 331

Qy     515 CAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATA-CAGACTGCAGGCAGACACCTCCCT 573
           || |   ||||||| |||||| || | || ||||  |||||||| |||||  || || |
Db     332 GAAAAGTCACGGGAAGTTCCTTCAGATTCCGGATATAAGACTGCAAGCAGAAGCCGCCTT 391

Qy     574 TCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGC 633
           |||| || |||||||||||| ||||| ||||| ||| |||| |  ||||||||||||||
Db     392 TCCTGGTGCTGCTGGTGACGTCATCTCACAAGCAGCTGGCCGCTCGCATGGCCATCCGCA 451

Qy     634 AGACGTGGGGGAAAGAGAGGATGGTGAAGGG-AAAGCAGCTGAAGACATTCTTCCTCCTG 692
           |||||||||| | |||||     ||| |||| | | ||| ||| ||| |||||||| |||
Db     452 AGACGTGGGGTAGAGAGACATCTGTGCAGGGCACAACAGGTGAGGACCTTCTTCCTTCTG 511

Qy     693 GGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGG 752
           |||||| ||  |||  | | ||| | | | |       |||  |     |    |     |
Db     512 GGGACCTCCGACAGCACCGAGGAGATGGACGCCACAAACCCTGGAGAGCGAGCAGCACCG 571

Qy     753 GACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATG 812
           |||||||||||||||| |||   || | |||      | ||||||||||||||||||||
Db     572 GACATTATCCAGAAGGACTTCAAGGATGCCTAACTTCAACTGACCCTGAAGACCATGATG 631

Qy     813 GGCAT 817
           || ||
Db     632 GGTAT 636
```

RESULT 10
AK003837
LOCUS AK003837 1514 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length enriched library, clone:1110020A06:UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 3, full insert sequence.
ACCESSION AK003837
VERSION AK003837.1 GI:12834746
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1110020A06.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5 (bases 1 to 1514)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1514
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1110020A06"
/db_xref="MGD:MGI:1898009"
/db_xref="taxon:10090"
/clone="1110020A06"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18 days embryo"
gene 1. .1514
/gene="B3galt3"
CDS <1. .769
/gene="B3galt3"
/note="UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 3
data source:MGD, source key:MGI:1349405, evidence:ISS putative"
/codon_start=2
/protein_id="BAB23028.1"
/db_xref="GI:12834747"
/db_xref="MGD:MGI:1349405"
/translation="NPFLVILVTSRPSDVKARQAIRVTWGEKKSWWGYEVLTPFLLGQQAEREDKTLALSLEDEHVLYGDIIRQDFLDTYNNLTLKTIMAFRWVMEFCPNAKYIMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFFHKNHISYQEYPFKVFPPYCSGLGYIMSGDLVPRVYEMMSHVKPIKFEDVYVGICLNLLKVDIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSKEIITFWQVMLRNTTCHY"
BASE COUNT 449 a 288 c 323 g 454 t
ORIGIN

Query Match 4.9%; Score 135; DB 11; Length 1514;
Best Local Similarity 51.0%; Pred. No. 8.1e-14;
Matches 374; Conservative 0; Mismatches 350; Indels 9; Gaps 2;

```
Qy    570 CCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATC 629
          || ||||| ||| | ||||||||||||  || |  | |  ||   |   |    |||||
Db      5 CCATTCCTGGTCATCCTGGTGACCTCACGCCCCTCAGATGTGAAAGCCAGACAAGCCATT 64

Qy    630 CGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTC 689
           |    || |||||  |  | | |     |  |||| |  || ||   || || ||| |
```

```
Db      65 AGAGTTACTTGGGGTGAGAAAAAGTCCTGGTGGGGATATGAGGTGCTCACGTTTTTCTTA 124

Qy     690 CTGGG---GACCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGA 746
           || ||   |    |    || | ||   ||||| ||   ||  |  ||||    |||
Db     125 CTAGGCCAGCAGGCTGAAAGGGAAGACAAAACGTTAGCCCTGTCCTTGGAGGATGAGCAC 184

Qy     747 C------ACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTG 800
             | || || ||||| | | | || || ||||||   ||| | ||  ||||| ||
Db     185 GTTCTCTATGGTGATATTATACGGCAAGACTTTCTAGACACATATAATAACTTGACCTTG 244

Qy     801 AAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATG 860
           || ||||| ||||    |    |||||       ||||| ||  | ||   || | | |||
Db     245 AAAACCATTATGGCTTTCAGGTGGGTAATGGAGTTTTGCCCCAATGCCAAGTATATCATG 304

Qy     861 AAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAA 920
           |||||||| | ||  | ||||||||   || | || |     |    ||| | ||   |
Db     305 AAAACAGACACTGATGTTTTCATCAACACTGGCAATTTAGTCAAGTATCTTTTAAACCTA 364

Qy     921 AACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAG 980
           |||    ||   | |||||||||| ||||       | |  || | | | ||   ||
Db     365 AACCACTCAGAGAAGTTTTTCACGGGCTATCCTCTAATTGATAACTATTCCTATAGAGGA 424

Qy     981 CCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 1040
           |||   ||        |    |    || || || |   |   || ||| || | |
Db     425 TTTTTCCATAAAAACCACATTTCATACCAAGAGTACCCCTTCAAGGTGTTCCCTCCCTAC 484

Qy    1041 TGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCC 1100
           |||  |||    |||||| |  | || ||||| ||| |     ||| ||| |  |
Db     485 TGCAGCGGGCTGGGCTACATTATGTCCGGCGACCTGGTGCCCAGGGTCTACGAGATGATG 544

Qy    1101 AAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTG 1160
           |    |||      ||| ||  | |||||||| | ||| ||  ||||  |  |   | |
Db     545 AGTCACGTGAAGCCCATCAAGTTTGAAGACGTTTATGTTGGCATCTGTTTGAATTTGTTA 604

Qy    1161 AACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTC 1220
           ||  |     |  |    |   |     || ||||      |  | | |||
Db     605 AAAGTGGACATTCATATTCCAGAAGACACAAACCTTTTCTTTCTGTACAGAATCCACTTG 664

Qy    1221 TCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTG 1280
           |||||  |   ||||  |  |  | ||   |||   | |     |   |    || |
Db     665 GATGTATGTCAGCTCAGACGCGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATC 724

Qy    1281 GACTACTGGCAGG 1293
           | ||||||||
Db     725 ACATTCTGGCAGG 737
```

```
RESULT 11
BQ021377/c
LOCUS       BQ021377                491 bp    mRNA    linear   EST 27-MAR-2002
DEFINITION  UI-H-DH1-axf-p-12-0-UI.s1 NCI_CGAP_DH1 Homo sapiens cDNA clone
            IMAGE:5828219 3', mRNA sequence.
ACCESSION   BQ021377
VERSION     BQ021377.1  GI:19756655
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            The following repetitive elements were found in this cDNA
            sequence: 32-98, >MLT2D#LTR/Retroviral (matched compliment) 32-76,
            >MLT2B2 (matched compliment) 137-323, >MLT2B2 (matched compliment)
            250-465, >MLT2B2 (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES            Location/Qualifiers
     source         1. .491
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5828219"
                    /clone_lib="NCI_CGAP_DH1"
                    /tissue_type="Metastatic Chondrosarcoma"
                    /dev_stage="Adult"
                    /lab_host="DH10B (Life Technologies)"
                    /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                    modified polylinker; Site_1: EcoR I; Site_2: Not I;
                    NCI_CGAP_DH1 is a normalized cDNA library containing the
                    following tissue(s): VS-8 Cell line from Metastatic
                    Chondrosarcoma in Lung. The library was constructed
                    according to Bonaldo, Lennon and Soares, Genome Research,
                    6:791-806, 1996. First strand cDNA synthesis was primed
                    with an oligo-dT primer containing a Not I site. Double
                    stranded cDNA was ligated to an EcoR I adaptor, digested
                    with Not I, and cloned directionally into pT7T3-Pac
                    vector. The oligonucleotide used to prime the synthesis of
                    first-strand cDNA contains a library tag sequence that is
                    located between the Not I site and the (dT)18 tail. The
                    sequence tag for this library is AGATCATTGC.
                    TAG_LIB=UI-H-DH1
                    TAG_TISSUE=lung
                    TAG_SEQ=AGATCATTGC"
BASE COUNT      144 a     93 c    123 g    131 t
ORIGIN
```

```
  Query Match             4.8%;  Score 133.6;  DB 14;  Length 491;
  Best Local Similarity   78.9%;  Pred. No. 2.3e-13;
  Matches  172;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    1;

Qy      12 TTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCAGAGCTGCAGGC 71
           | ||||  |||||||||| || | |||||||||| ||||||||||||||||||| || | |||
Db     420 TCTCTCCTCTGGAGCTGAGACACTCTTCTTCTTCTGCCCTTGGACATCAGAACTCCTGGC 361

Qy      72 TCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGGCCTTTG 131
           |||| |||||||| ||   |||| || ||||| |   |   ||||||||||||||||||||
Db     360 TCTCCGGCCTTTGAACTTCAGGACTTGTACCAGG--AGGCCCTGGGTTCTCAGGCCTTTG 303

Qy     132 GCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCTTGGACT 191
           || |||||||| |||||||| ||  || |  ||||||||||||||| |  | |||  |||
Db     302 GCTTTGGACTGAGAGTTACACAATCAGCTTCCCTGGTTCTGAGGCTTTCAGACTTAAACT 243

Qy     192 GAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCA 229
           ||||||  || |  ||||||||| ||||||||| | ||
Db     242 GAGCCATGCTACCAGCATCCCAGGGTCTCCAGCCTACA 205
```

```
RESULT 12
BC013258
LOCUS       BC013258                3199 bp    mRNA    linear   HTC 27-AUG-2001
DEFINITION  Homo sapiens, clone IMAGE:3867325, mRNA.
ACCESSION   BC013258
VERSION     BC013258.1  GI:15301554
KEYWORDS    HTC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3199)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
```

```
  REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Baylor College of Medicine Human Genome
             Sequencing Center
             Center code: BCM-HGSC
             Web site: http://www.hgsc.bcm.tmc.edu/cdna/
             Contact: villalon@bcm.tmc.edu.
             Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
             A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
             Muzny,D.M., Gibbs,R.A.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 13 Row: f Column: 3
             This clone was selected for full length sequencing because it
             passed the following selection criteria: Hexamer frequency ORF
             analysis
             This clone has the following problem: no 5' EST match.
FEATURES               Location/Qualifiers
     source            1. .3199
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="IMAGE:3867325"
                       /tissue_type="Eye, retinoblastoma"
                       /clone_lib="NIH_MGC_67"
                       /lab_host="DH10B"
                       /note="Vector: pCMV-SPORT6"
BASE COUNT       841 a    747 c    771 g    840 t
ORIGIN

  Query Match             4.8%;  Score 133.6;  DB 11;  Length 3199;
  Best Local Similarity   78.9%;  Pred. No. 1e-13;
  Matches  172;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    1;

Qy      12 TTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCAGAGCTGCAGGC 71
           | ||||  |||||||||| || | |||||||||| ||||||||||||||||||| || | |||
Db    2785 TCTCTCCTCTGGAGCTGAGACACTCTTCTTCTTCTGCCCTTGGACATCAGAACTCCTGGC 2844

Qy      72 TCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGGCCTTTG 131
           |||| |||||||| ||   |||| || ||||| |   |   ||||||||||||||||||
Db    2845 TCTCCGGCCTTTGAACTTCAGGACTTGTACCAGG--AGGCCCTGGGTTCTCAGGCCTTTG 2902

Qy     132 GCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCTTGGACT 191
           || |||||||||| |||||||| ||  || |  |||||||||||||| |  | |||  |||
Db    2903 GCTTTGGACTGAGAGTTACACAATCAGCTTCCCTGGTTCTGAGGCTTTCAGACTTAAACT 2962

Qy     192 GAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCA 229
           ||||||  || |  |||||||||| |||||||||  | ||
Db    2963 GAGCCATGCTACCAGCATCCCAGGGTCTCCAGCCTACA 3000


RESULT 13
BJ501651
LOCUS        BJ501651                596 bp    mRNA    linear   EST 08-AUG-2002
DEFINITION   BJ501651 MF01FSA cDNA Oryzias latipes cDNA clone MF01FSA049L05 5',
             mRNA sequence.
ACCESSION    BJ501651
VERSION      BJ501651.1  GI:22153613
KEYWORDS     EST.
SOURCE       Japanese medaka.
  ORGANISM   Oryzias latipes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
             Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE    1  (bases 1 to 596)
  AUTHORS    Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE      Medaka EST Project in Takeda's lab
  JOURNAL    Unpublished (2001)
COMMENT      Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES               Location/Qualifiers
     source            1. .596
                       /organism="Oryzias latipes"
                       /strain="d-rR"
                       /db_xref="taxon:8090"
                       /clone="MF01FSA049L05"
                       /clone_lib="MF01FSA cDNA"
                       /sex="mixture of female and male"
                       /tissue_type="whole embryo"
                       /dev_stage="fry stage 40"
BASE COUNT       138 a    168 c    151 g    137 t      2 others
ORIGIN

  Query Match             4.8%;  Score 132.2;  DB 13;  Length 596;
  Best Local Similarity   55.2%;  Pred. No. 3.7e-13;
  Matches  277;  Conservative    0;  Mismatches  222;  Indels    3;  Gaps    1;

Qy     668 GCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGT 727
           ||   ||   ||  | ||| |  |||||  || ||     || |     |   |  |||
Db      13 GCGAGTGGTCACCCTTTTCTTGTTGGGGCGCAGCATGGACGCTGTCCTGAACCAGATGGT 72

Qy     728 GGACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTA 787
           ||| |||||||| |||          ||  |  ||  | |||| ||  | |||   ||  |
Db      73 GGAGCAGGAGAGTCAGATCTTTCATGATGTAGTCGTGGAGGACTTTATTGACTCGTACCA 132

Qy     788 CAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGC 847
           ||| ||||| || |||||| ||||||| ||    |||||   |  ||  ||  |  ||||
Db     133 CAACCTGACGCTNAAGACCCTGATGGGTATGCGCTGGGTGGCTACCTACTGCACCAAGGC 192

Qy     848 GGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACT 907
           | |||  | ||  || |||   ||||| ||  ||||  | ||  | ||  |   |
Db     193 CCAATATGTACTCAAGACGGACAGTGACATCTTTGTCAACATGGAGAACCTCATCTACAA 252

Qy     908 GCTTCTGAAG---AAAAACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGA 964
           |||||||||   |  | ||  | |  ||    |    |      |      |||| |
Db     253 CCTTCTGAAGCCCACCACCAAGCCTAGGAGAAGGTATTTTACAGGTTACGTCATCAACGG 312

Qy     965 GTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGA 1024
           |   || |||||  |   |  ||||||||||    |       |   |||||
Db     313 GGGGCCGATCAGAGACATACGCAGCAAGTGGTACATGCCCCGGGATCTTTATCCCGACAG 372

Qy    1025 CAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCA 1084
           ||  |||||||| ||||||||||||||||| ||||| || || |  || || ||    |
Db     373 CAAATACCCACCCTTCTGCTCCGGCACCGGGTACGTCTTCTCAGCAGATGTAGCTGAGCT 432

Qy    1085 GGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCT 1144
           | ||| |  ||||   |  |  |     | |  | ||||| |||||||| ||| || |
Db     433 CATATACNAGATCTCTTTGCACACCAGGCTCCTCCACCTGGAGGACGTGTATGTTGGAGT 492

Qy    1145 CTGCCTCGAAAGGCTGAACATC 1166
           ||||||   | ||||  ||||
Db     493 TTGCCTCCGTAAGCTGGGCATC 514


RESULT 14
BC028571
LOCUS        BC028571                3564 bp    mRNA    linear   HTC 21-AUG-2002
DEFINITION   Homo sapiens, UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase,
             polypeptide 3, clone IMAGE:4838965, mRNA.
ACCESSION    BC028571
VERSION      BC028571.1  GI:22382221
KEYWORDS     HTC.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 3564)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 34 Row: p Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 15451873
            This clone has the following problem: incomplete processing.
FEATURES             Location/Qualifiers
     source          1. .3564
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4838965"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
BASE COUNT     1048 a    661 c    741 g   1114 t
ORIGIN

  Query Match             4.6%;  Score 128.8;  DB 11;  Length 3564;
  Best Local Similarity   48.7%;  Pred. No. 6.4e-13;
  Matches  386;  Conservative    0;  Mismatches  397;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||  |||||  |||||  |     || || |||||  ||  |  ||
Db     864 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 923

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
             || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db     924 TATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 983

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | | |      |  |||| |  || |    |||||  ||| |
Db     984 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 1043

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                |  || |  |    ||         | ||  | ||   ||
Db    1044 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 1103

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
             |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db    1104 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 1163

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    1164 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 1223

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | ||||||||||  || | || |     |    ||| | ||   ||
Db    1224 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 1283

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||| || |       | |  || | | | ||   ||
Db    1284 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 1343

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
           | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db    1344 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1403

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1404 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1463

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| || | ||||| || | ||| ||| |||| |  |     | |
Db    1464 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1523

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |    |          || |||| |    |  | |  ||
Db    1524 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1583

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
           || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1584 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1643

Qy    1282 ACTACTGGCAGG 1293
           |  |||||||
Db    1644 CTTTTTGGCAGG 1655


RESULT 15
BB595953
LOCUS       BB595953                 233 bp    mRNA    linear   EST 30-NOV-2000
DEFINITION  BB595953 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C230001K09 5', mRNA sequence.
ACCESSION   BB595953
VERSION     BB595953.1  GI:11492555
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
            ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
            Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
            Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
            Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
            ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
```

```
          further details.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C230001K09"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 479.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       57 a     57 c     74 g     44 t      1 others
ORIGIN

  Query Match             4.6%;  Score 127;  DB 10;  Length 233;
  Best Local Similarity   71.6%;  Pred. No. 4.5e-12;
  Matches  166;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     570 CCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATC 629
           || ||||| || ||| ||||||| ||||| ||||| ||| |||  |  ||||||| ||||
Db       2 CCTTTCCTGGTGCTGGTGGTGACGTCATCTCACAAGCAGCTGGGCGCTCGCATGGGCATC 61

Qy     630 CGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTC 689
           ||| ||| |||||| |||||||     ||| ||||  | ||| ||| ||| | ||||||
Db      62 CGGAAGAAGTGGGGTAAAGAGACATCTGTGGAGGGCCAACAGGTGAGGACCTCCTTCCTT 121

Qy     690 CTGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACAC 749
           ||||||| | ||  |||  | | ||||| | | |       |  ||| ||| ||    ||
Db     122 CTGGGGATCTCCGACAGGACCGAGGAAATGGACGCCACAACCGTGGAAAGCGAGNAGTAC 181

Qy     750 GGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
           ||||| ||||||||||||||||||   || | ||| | |||  |||||||||
Db     182 GGGGATATTATCCAGAAGGATTTCAAGGATGCCTACTCCAACGTGACCCTGA 233


Search completed: April 11, 2003, 22:37:14
Job time : 3113.38 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 15:37:18 ; Search time 76.6272 Seconds
(without alignments)
11106.080 Million cell updates/sec

Title: US-09-914-152-2
Perfect score: 2775
Sequence: 1 gtgaattcctctttctctgc.....aaaaaaaaaaaaaaaaaaa 2775

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	133.6	4.8	3502	2	US-08-724-394A-16	Sequence 16, Appl
2	124.8	4.5	333	4	US-08-905-223-218	Sequence 218, App
3	114.8	4.1	1134	4	US-09-482-180A-3	Sequence 3, Appli
4	78.4	2.8	1420	4	US-09-482-180A-1	Sequence 1, Appli
5	68.6	2.5	1191	4	US-09-459-133-3	Sequence 3, Appli
6	67	2.4	1167	4	US-09-459-133-14	Sequence 14, Appl
7	61	2.2	2550	6	5258287-23	Patent No. 5258287
c 8	58.2	2.1	703	4	US-09-313-300-6	Sequence 6, Appli
9	57	2.1	1129	4	US-09-227-357-40	Sequence 40, Appl
10	57	2.1	1325	1	US-08-306-691B-51	Sequence 51, Appl
11	56.6	2.0	1441	4	US-08-821-994-63	Sequence 63, Appl
12	56.4	2.0	1699	4	US-09-152-060-19	Sequence 19, Appl
13	56.4	2.0	2239	4	US-09-196-390-1	Sequence 1, Appli
14	56.2	2.0	577	4	US-09-149-476-313	Sequence 313, App
15	55.4	2.0	1325	2	US-08-464-517-1	Sequence 1, Appli
16	55.4	2.0	1325	2	US-08-246-361A-1	Sequence 1, Appli
17	55.4	2.0	1325	3	US-08-463-772-1	Sequence 1, Appli
18	55.4	2.0	1325	5	PCT-US93-05000-1	Sequence 1, Appli
19	55.2	2.0	2837	2	US-08-993-228-11	Sequence 11, Appl
20	54.4	2.0	1434	2	US-09-055-097-2	Sequence 2, Appli
21	54.2	2.0	458	1	US-08-524-757-1	Sequence 1, Appli
22	54.2	2.0	2246	4	US-09-363-708-3	Sequence 3, Appli
23	54.2	2.0	2447	2	US-09-014-969-14	Sequence 14, Appl
24	54.2	2.0	4843	3	US-08-986-485-1	Sequence 1, Appli
25	54	1.9	1825	4	US-09-461-697-75	Sequence 75, Appl
26	54	1.9	50000	4	US-09-146-053-4	Sequence 4, Appli
27	53.8	1.9	3214	1	US-08-484-105-17	Sequence 17, Appl
28	53.8	1.9	3214	1	US-08-484-106-17	Sequence 17, Appl
29	53.8	1.9	3300	4	US-09-336-643A-82	Sequence 82, Appl
30	53.8	1.9	6671	1	US-08-280-443-1	Sequence 1, Appli
31	53.8	1.9	6671	1	US-08-457-459-1	Sequence 1, Appli
32	53.8	1.9	6671	1	US-08-555-678-1	Sequence 1, Appli
33	53.8	1.9	6671	5	PCT-US95-02275-1	Sequence 1, Appli
34	53.4	1.9	1364	1	US-08-265-087-3	Sequence 3, Appli
35	53.4	1.9	1364	1	US-08-621-493-3	Sequence 3, Appli
36	53.4	1.9	1364	2	US-08-965-688-3	Sequence 3, Appli
37	53.4	1.9	1364	4	US-09-260-173-3	Sequence 3, Appli
38	53.4	1.9	2806	4	US-09-653-839-9	Sequence 9, Appli
39	53.4	1.9	3080	4	US-09-099-041A-25	Sequence 25, Appl
40	53.4	1.9	3080	4	US-09-245-281-25	Sequence 25, Appl
41	53.4	1.9	3080	4	US-09-207-359B-25	Sequence 25, Appl
42	53.2	1.9	1098	3	US-09-248-335-35	Sequence 35, Appl
43	53.2	1.9	3410	4	US-09-020-956-110	Sequence 110, App
44	53.2	1.9	3410	4	US-09-030-607-110	Sequence 110, App
45	53.2	1.9	3410	4	US-09-605-785-110	Sequence 110, App

ALIGNMENTS

```
RESULT 1
US-08-724-394A-16
; Sequence 16, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237e1
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3502 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..3502
;      OTHER INFORMATION:  /note= "cDNA 32"
US-08-724-394A-16
```

```
  Query Match             4.8%;  Score 133.6;  DB 2;  Length 3502;
  Best Local Similarity   78.9%;  Pred. No. 3.6e-27;
  Matches  172;  Conservative    0;  Mismatches   44;  Indels    2;  Gaps    1;

Qy      12 TTTCTCTGCTGGAGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCAGAGCTGCAGGC 71
           | ||||  |||||||||| || | |||||||||| ||||||||||||||||||| || | |||
Db    3101 TCTCTCCTCTGGAGCTGAGACACTCTTCTTCTTCTGCCCTTGGACATCAGAACTCCTGGC 3160

Qy      72 TCTCTGGCCTTTGGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGGCCTTTG 131
           |||| |||||||| ||   |||| || ||||| |   |   ||||||||||||||||||
Db    3161 TCTCCGGCCTTTGAACTTCAGGACTTGTACCAGG--AGGCCCTGGGTTCTCAGGCCTTTG 3218

Qy     132 GCCTTGGACTGATAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCTTGGACT 191
           || ||||||||| |||||||| ||  || |  |||||||||||||| |  | |||  |||
Db    3219 GCTTTGGACTGAGAGTTACACAATCAGCTTCCCTGGTTCTGAGGCTTTCAGACTTAAACT 3278

Qy     192 GAGCCACACTCCTGGCATCCCAGCGTCTCCAGCTTGCA 229
           ||||||  || |  |||||||| |||||||| | ||
Db    3279 GAGCCATGCTACCAGCATCCCAGGGTCTCCAGCCTACA 3316


RESULT 2
US-08-905-223-218
; Sequence 218, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Lacroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  218:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  333 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Brain
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  32..331
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 99
;      OTHER INFORMATION:  region 1..300
;      OTHER INFORMATION:  id R13004
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  114..274
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 99
;      OTHER INFORMATION:  region 54..214
;      OTHER INFORMATION:  id T80337
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  272..331
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 213..272
;      OTHER INFORMATION:  id T80337
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  66..106
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 6..46
;      OTHER INFORMATION:  id T80337
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  101..278
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 94
;      OTHER INFORMATION:  region 70..247
;      OTHER INFORMATION:  id T08840
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  33..113
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 98
;      OTHER INFORMATION:  region 1..81
;      OTHER INFORMATION:  id T08840
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  101..249
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 95
;      OTHER INFORMATION:  region 72..220
;      OTHER INFORMATION:  id HSC0CF041
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  31..112
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 1..82
;      OTHER INFORMATION:  id HSC0CF041
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  247..321
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 5.8
;      OTHER INFORMATION:  seq LWISACAMLLCHG/SL
US-08-905-223-218

  Query Match             4.5%;  Score 124.8;  DB 4;  Length 333;
  Best Local Similarity   74.2%;  Pred. No. 2.5e-25;
  Matches  184;  Conservative    0;  Mismatches   62;  Indels    2;  Gaps    2;

Qy      24 AGCTGGGATATTCTTCTTCTCCTGCCCTTGGACATCAGAGCTGCAGGCTCTCTGGCCTTT 83
           |||| ||| |||||||| |||||| || |||||||||| |  |||||||||  ||||||
Db       2 AGCTAGGACATTCTTCTCCTCCTGGCCCTGGACATCAGAACCCCAGGCTCTCCAGCCTTT 61

Qy      84 GGACCCGAGGATTTATACCAAGCAGGTTTCTGGGTTCTCAGGCCTTTGGCCTTGGACTGA 143
```

```
Db     62 GGACTTCAGGACTGACA-CAAGCAACCTGCTGGGTTCTTAGGCCTTTGG-CTTGTACTGA 119

Qy    144 TAGTTACACCATTGGCATATCTGGTTCTGAGGCTCTTGGTCTTGGACTGAGCCACACTCC 203

Db    120 GACTTACACCATCAGCTTCCCTGGTCCTGAGACTTTTGGACTTGGATTGAGCCACGCTAC 179

Qy    204 TGGCATCCCAGCGTCTCCAGCTTGCATGGCCTGTCACGTGATTCCTGTCAGAATCACCAT 263

Db    180 TGGCATCCCAGGATCTCCAGCTTGCAGACAGCCTGTCGTGGGACTTCACAGCCTCCATAA 239

Qy    264 TTTTGGTA 271

Db    240 TTATAGAA 247


RESULT 3
US-09-482-180A-3
; Sequence 3, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Gao, Zeren
;  APPLICANT:  Whitmore, Theodore E.
;  APPLICANT:  Jaspers, Stephen
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
;  TITLE OF INVENTION:  BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZNSSP6
;  FILE REFERENCE: 98-80
;  CURRENT APPLICATION NUMBER: US/09/482,180A
;  CURRENT FILING DATE:  2000-01-12
;  PRIOR APPLICATION NUMBER: 60/115,721
;  PRIOR FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1134
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: degenerate sequence
;   NAME/KEY: variation
;   LOCATION: (1)...(1134)
;   OTHER INFORMATION: n is any nucleotide
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1134)
;   OTHER INFORMATION: n = A,T,C or G
US-09-482-180A-3

  Query Match             4.1%;  Score 114.8;  DB 4;  Length 1134;
  Best Local Similarity   31.3%;  Pred. No. 3.3e-22;
  Matches  209;  Conservative   98;  Mismatches  358;  Indels    3;  Gaps    1;

Qy    504 TCCTTTGTTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAG 563

Db    286 TTYYTNACNTAYMGNCAYTGYMGNAAYTTYWSNATHYTNYTNGARCCNWSNGGNTGYWSN 345

Qy    564 ACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATG 623

Db    346 AARGAYACNTTYYTNYTNYTNGCNATHAARWSNCARCCNGGNCAYGTNGARMGNMGNGCN 405

Qy    624 GCCATCCGGCAGACGTGGGGGA---AAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACA 680

Db    406 GCNATHMGNWSNACNTGGGGNMGNGTNGGNGGNTGGGCNMGNGGNMGNCARYTNAARYTN 465

Qy    681 TTCTTCCTCCTGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGC 740

Db    466 GTNTTYYTNYTNGGNGTNGCNGGNWSNGCNCCNCCNGCNCARYTNYTNGCNTAYGARWSN 525

Qy    741 CAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTG 800

Db    526 MGNGARTTYGAYGAYATHYTNCARTGGGAYTTYACNGARGAYTTYTTYAAYYTNACNYTN 585

Qy    801 AAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATG 860

Db    586 AARGARYTNCAYYTNCARMGNTGGGTNGTNGCNGCNTGYCCNCARGCNCAYTTYATGYTN 645

Qy    861 AAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAA 920

Db    646 AARGGNGAYGAYGAYGTNTTYGTNCAYGTNCCNAAYGTNYTNGARTTYYTNGAYGGNTGG 705

Qy    921 AACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAG 980

Db    706 GAYCCNGCNCARGAYYTNYTNGTNGGNGAYGTNATHMGNCARGCNYTNCCNAAYMGNAAY 765

Qy    981 CCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTC 1040

Db    766 ACNAARGTNAARTAYTTYATHCCNCCNWSNATGTAYMGNGCNACNCAYTAYCCNCCNTAY 825

Qy   1041 TGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCC 1100

Db    826 GCNGGNGGNGGNGGNTAYGTNATGWSNMGNGCNACNGTNMGNMGNYTNCARGCNATHATG 885

Qy   1101 AAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTG 1160

Db    886 GARGAYGCNGARYTNYTNWSNATHGAYGAYGTNTTYGTNGGNATGTGYYTNMGNMGNYTN 945

Qy   1161 AACATCAG 1168

Db    946 GGNYTNWS 953


RESULT 4
US-09-482-180A-1
; Sequence 1, Application US/09482180A
; Patent No. 6361985
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Gao, Zeren
;  APPLICANT:  Whitmore, Theodore E.
;  APPLICANT:  Jaspers, Stephen
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING
;  TITLE OF INVENTION:  BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOG, ZNSSP6
;  FILE REFERENCE: 98-80
;  CURRENT APPLICATION NUMBER: US/09/482,180A
;  CURRENT FILING DATE:  2000-01-12
;  PRIOR APPLICATION NUMBER: 60/115,721
;  PRIOR FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1420
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (135)...(1271)
US-09-482-180A-1

  Query Match             2.8%;  Score 78.4;  DB 4;  Length 1420;
  Best Local Similarity   46.1%;  Pred. No. 6e-12;
  Matches  299;  Conservative    0;  Mismatches  346;  Indels    3;  Gaps    1;

Qy    524 CGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCT 583

Db    440 CCGAAATTTCTCTATCTTGCTGGAGCCTTCAGGCTGTTCCAAGGATACCTTCTTGCTCCT 499

Qy    584 GCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGG 643

Db    500 GGCCATCAAGTCACAGCCTGGTCACGTGGAGCGACGTGCGGCTATCCGCAGCACGTGGGG 559

Qy    644 GAAAGAGAGG---ATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCAC 700

Db    560 CAGGGTGGGGGGATGGGCTAGGGGCCGGCAGCTGAAGCTGGTGTTCCTCCTAGGGGTGGC 619
```

```
Qy    701 CAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTAT 760
Db    620 AGGATCCGCTCCCCCAGCCCAGCTGCTGGCCTATGAGAGTAGGGAGTTTGATGACATCCT 679

Qy    761 CCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGA 820
Db    680 CCAGTGGGACTTCACTGAGGACTTCTTCAACCTGACGCTCAAGGAGCTGCACCTGCAGCG 739

Qy    821 ATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTT 880
Db    740 CTGGGTGGTGGCTGCCTGCCCCCAGGCCCATTTCATGCTAAAGGGAGATGACGATGTCTT 799

Qy    881 CATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTT 940
Db    800 TGTCCACGTCCCCAACGTGTTAGAGTTCCTGGATGGCTGGGACCCAGCCCAGGACCTCCT 859

Qy    941 CACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGT 1000
Db    860 GGTGGGAGATGTCATCCGCCAAGCCCTGCCCAACAGGAACACTAAGGTCAAATACTTCAT 919

Qy   1001 CAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGT 1060
Db    920 CCCACCCTCAATGTACAGGGCCACCCACTACCCACCCTATGCTGGTGGGGGAGGATATGT 979

Qy   1061 GTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAA 1120
Db    980 CATGTCCAGAGCCACAGTCGCGCGCCTCCAGGCTATCATGGAAGATGCTGAACTCCTCTC 1039

Qy   1121 ACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAG 1168
Db   1040 CATTGATGATGTCTTTGTGGGTATGTGCCTGAGGAGGCTGGGGCTGAG 1087


RESULT 5
US-09-459-133-3
; Sequence 3, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Jaspers, Stephen R.
;  APPLICANT:  Gao, Zeren
;  TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
;  FILE REFERENCE: 98-77
;  CURRENT APPLICATION NUMBER: US/09/459,133
;  CURRENT FILING DATE:  1999-12-10
;  PRIOR APPLICATION NUMBER: 60/111,697
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1191
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1191)
;   OTHER INFORMATION: n = A,T,C or G
US-09-459-133-3

  Query Match             2.5%;  Score 68.6;  DB 4;  Length 1191;
  Best Local Similarity   28.3%;  Pred. No. 3e-09;
  Matches  211;  Conservative   95;  Mismatches  428;  Indels   12;  Gaps    3;

Qy    570 CCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATC 629
Db    442 CCNTAYYTNYTNYTNGCNGTNAARWSNGARCCNGGNMGNTTYGCNGARMGNCARGCNGTN 501

Qy    630 CGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTC 689
Db    502 MGNGARACNTGGGGNWSNCCNGCNCCNGGNATHMGNYTNYTNTTYYTNYTNGGNWSNCCN 561

Qy    690 CTGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACAC 749
Db    562 GTNGGNGARGCNGGNCCNGAYYTNGAYWS---NYTNGTNGCNTGGGARWSNMGNMGNTAY 618

Qy    750 GGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATG 809
Db    619 WSNGAYYTNYTNYTNTGGGAYTTYYTNGAYGTNCCNTTYAAYCARACNYTNAARGAYYTN 678

Qy    810 ATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGAC 869
Db    679 YTNYTNYTNGCNTGGYTNGGNMGNCAYTGYCCNACNGTNWSNTTYGTNYTNMGNGCNCAR 738

Qy    870 TCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCT------TCTGAAGAAAAAC 923
Db    739 GAYGAYGCNTTYGTNCAYACNCCNGCNYTNYTNGCNCAYYTNMGNGCNYTNCCNCCNGCN 798

Qy    924 AGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCA 983
Db    799 WSNGCNMGNWSNYTNTAYYTNGGNGARGTNTTYACNCARGCNATGCCNYTNMGNAARCCN 858

Qy    984 TTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGC 1043
Db    859 GGNGGNCCNTTYTAYGTNCCNGARWSNTTYTTYGARGGNG---GNTAYCCNGCNTAYGCN 915

Qy   1044 TCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAG 1103
Db    916 WSNGGNGGNGGNTAYGTNATHGCNGGNMGNYTNGCNCCNTGGYTNYTNMGNGCNGCNGCN 975

Qy   1104 AGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAAC 1163
Db    976 MGNGTNGCNCCNTTYCCNTTYGARGAYGTNTAYACNGGNYTNTGYATHMGNGCNYTNGGN 1035

Qy   1164 ATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCC 1223
Db   1036 YTNGTNCCNCARGCNCAYCCNGGNTTYYTNACNGCNTGGCCNGCNGAYMGNACNGCNGAY 1095

Qy   1224 GTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGAC 1283
Db   1096 CAYTGYGCNTTYMGNAAYYTNYTNYTNGTNMGNCCNYTNGGNCCNCARGCNWSNATHMGN 1155

Qy   1284 TACTGGCAGGCTCTAGAGAATTCCCG 1309
Db   1156 YTNTGGAARCARYTNCARGAYCCNMG 1181


RESULT 6
US-09-459-133-14
; Sequence 14, Application US/09459133
; Patent No. 6416988
; GENERAL INFORMATION:
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT:  Yamamoto, Gayle
;  APPLICANT:  Jaspers, Stephen R.
;  APPLICANT:  Gao, Zeren
;  TITLE OF INVENTION: BETA-1,3-GALACTOSYLTRANSFERASE HOMOLOGS
;  FILE REFERENCE: 98-77
;  CURRENT APPLICATION NUMBER: US/09/459,133
;  CURRENT FILING DATE:  1999-12-10
;  PRIOR APPLICATION NUMBER: 60/111,697
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
;   LENGTH: 1167
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: degenerate sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1167)
;   OTHER INFORMATION: n = A,T,C or G
```

US-09-459-133-14

Query Match 2.4%; Score 67; DB 4; Length 1167;
Best Local Similarity 31.1%; Pred. No. 8.3e-09;
Matches 183; Conservative 74; Mismatches 319; Indels 12; Gaps 3;

```
Qy    570 CCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATC 629
Db    418 CCNTAYYTNYTNYTNGCNGTNAARWSNGARCCNGGNCAYTTYGCNGCNMGNCARGCNGTN 477

Qy    630 CGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTC 689
Db    478 MGNGARACNTGGGGNWSNCCNGTNGCNGGNACNMGNYTNYTNTTYYTNYTNGGNWSNCCN 537

Qy    690 CTGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACAC 749
Db    538 YTNGG---NATGGGNGGNCCNGAYYTNMGNWSNYTNGTNACNTGGGARWSNMGNMGNTAY 594

Qy    750 GGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATG 809
Db    595 GGNGAYYTNYTNYTNTGGGAYTTYYTNGAYGTNCCNTAYAAYMGNACNYTNAARGAYYTN 654

Qy    810 ATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGAC 869
Db    655 YTNYTNYTNACNTGGYTNWSNCAYCAYTGYCCNGAYGTNAAYTTYGTNYTNCARGTNCAR 714

Qy    870 TCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAA------GAAAAAC 923
Db    715 GAYGAYGCNTTYGTNCAYATHCCNGCNYTNYTNGARCAYYTNCARACNYTNCCNCCNACN 774

Qy    924 AGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCA 983
Db    775 TGGGCNMGNWSNYTNTAYYTNGGNGARATHTTYACNCARGCNAARCCNYTNMGNAARCCN 834

Qy    984 TTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGC 1043
Db    835 GGNGGNCCNTTYTAYGTNCCNAARACNTTYTTYGARGGNGAY---TAYCCNGCNTAYGCN 891

Qy   1044 TCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAG 1103
Db    892 WSNGGNGGNGGNTAYGTNATHWSNGGNMGNYTNGCNCCNTGGYTNYTNCARGCNGCNGCN 951

Qy   1104 AGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTC 1151
Db    952 MGNGTNGCNCCNTTYCCNTTYGAYGAYGTNTAYACNGGNTTYTGYTTY 999
```

RESULT 7
5258287-23
;Patent No. 5258287
; APPLICANT: BAXTER, ROBERT C.;WOOD, WILLIAM I.
; TITLE OF INVENTION: DNA ENCODING AND METHODS OF PRODUCTION
;OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN BP53
; NUMBER OF SEQUENCES: 58
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/171,623
; FILING DATE: 22-MAR-1988
;SEQ ID NO:23:
; LENGTH: 2550
5258287-23

Query Match 2.2%; Score 61; DB 6; Length 2550;
Best Local Similarity 64.5%; Pred. No. 6.5e-07;
Matches 91; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy   2635 TTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACA 2694
Db   2381 TTTTCTTTATTTTTTTAACTTTTTGGGGGGAAAAGTATTTTTGAGAAGTTTGTCTTGCAA 2440

Qy   2695 TGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAAAAAAA 2754
Db   2441 TGTATTTATAAATAGTAAATAAAGTTTTTACCATTAAAAAAAAAAAAAAAAAAAAAAAAA 2500

Qy   2755 AAAAAAAAAAAAAAAAAAAAA 2775
Db   2501 AAAAAAAAAAAAAAAAAAAAA 2521
```

RESULT 8
US-09-313-300-6/c
; Sequence 6, Application US/09313300
; Patent No. 6222027
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew, R.
; APPLICANT: Lal, Preeti
; APPLICANT: Yue, Henry
; APPLICANT: Tang, Tom, Y.
; APPLICANT: Baughn, Mariah, R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: MOLECULES EXPRESSED IN HIPPOCAMPUS
; FILE REFERENCE: PB-0012 US
; CURRENT APPLICATION NUMBER: US/09/313,300
; CURRENT FILING DATE: 1999-05-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY:
; OTHER INFORMATION: 1941247
; PUBLICATION INFORMATION:
US-09-313-300-6

Query Match 2.1%; Score 58.2; DB 4; Length 703;
Best Local Similarity 61.6%; Pred. No. 1.8e-06;
Matches 93; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy   2622 TTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAG 2681
Db    162 TTATAGTATGAAATTTCAATTTTACTTTGTACTGTACATCTTTTTACTTTAGAATTTGCA 103

Qy   2682 AAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAA 2741
Db    102 ATAAAGGGTTACGTCAATCTTGTTTTCAACTCTCAAAAAAAAAAAAAAAAAAAAAAAA 43

Qy   2742 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2772
Db     42 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 12
```

RESULT 9
US-09-227-357-40
; Sequence 40, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: PZ010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08

```
;  EARLIER APPLICATION NUMBER: 60/051,931
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,932
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,916
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,930
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,918
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,920
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,733
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,795
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,919
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,928
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/055,722
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,723
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,948
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,949
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,953
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,950
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,947
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,964
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/056,360
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,684
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,984
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,954
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/058,785
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,664
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,660
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,661
;  EARLIER FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 672
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
;   LENGTH: 1129
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1053)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-40

  Query Match             2.1%;  Score 57;  DB 4;  Length 1129;
  Best Local Similarity   60.4%;  Pred. No. 5.2e-06;
  Matches   93;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   2622 TTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAG 2681
          ||  ||  ||  |||| |  ||| || ||| | | |||  |    |  || |   | |
Db    953 TTTAAGGTATATTTTTCCCATTATTTATTTTCCTGACCTTAAAACAGCTTTTCTACTAAA 1012

Qy   2682 AAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAA 2741
          ||  ||      |||        | || |     ||   | |||     ||||||||||
Db   1013 AAATGGTGAGCAATGAAGACAATAAATTTTTCATTTTTCCNAAAAAAAAAAAAAAAAAAA 1072

Qy   2742 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2775
          ||||||||||||||||||||||||||||||||||
Db   1073 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1106


RESULT 10
US-08-306-691B-51
; Sequence 51, Application US/08306691B
; Patent No. 5734039
;  GENERAL INFORMATION:
;    APPLICANT:  Calabretta, Bruno
;    APPLICANT:  Skorski, Tomasz
;    TITLE OF INVENTION:  ANTISENSE
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;      STREET:  Two Penn Center, Suite 1800
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/306,691B
;      FILING DATE:  September 15, 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A.
;      REGISTRATION NUMBER:  30,480
;      REFERENCE/DOCKET NUMBER:  8321-8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-8383
;      TELEFAX:  (215) 568-5549
;      TELEX:  No. 5734039e
;  INFORMATION FOR SEQ ID NO:  51:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-306-691B-51

  Query Match             2.1%;  Score 57;  DB 1;  Length 1325;
  Best Local Similarity   62.1%;  Pred. No. 5.7e-06;
  Matches   90;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy   2631 TGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCT 2690
          || |||||| |||  | |||  |  | |||    | ||||       ||| ||
Db   1167 TGTTTTTTCCTTTGCTCTTTCCCCCTTCCATCTCTGACTTAAGCAAAAGAAAAAGATTAC 1226

Qy   2691 CACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAA 2750
          |  |   |||| |   | | | |    |   |||     ||||||||||||||||||
Db   1227 CCAAAAACTGTCTTTAAAAGAGAGAGAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1286

Qy   2751 AAAAAAAAAAAAAAAAAAAAAAAAA 2775
          |||||||||||||||||||||||||
Db   1287 AAAAAAAAAAAAAAAAAAAAAAAAA 1311


RESULT 11
US-08-821-994-63
; Sequence 63, Application US/08821994A
```

```
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-63

  Query Match             2.0%;  Score 56.6;  DB 4;  Length 1441;
  Best Local Similarity   59.7%;  Pred. No. 7.8e-06;
  Matches   95;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    2616 TCCTAATTATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTAC 2675
           ||   ||  | |  |  ||  ||   || |     || ||     | ||| | | ||||
Db    1283 TCTCTATGTTGGTCAATATCATTTCATTCAAAGAATATTTGCTTTGGCTTGATTATGTAT 1342

Qy    2676 TCAGAGAAGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAA 2735
           | ||||||      | | |  | | | |  |  | |||      |  |||     |||||
Db    1343 TAAGAGAAATATAATAAAAATGATATATTTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAA 1402

Qy    2736 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2774
           ||||||||||||||||||||||||||||||||||||||||
Db    1403 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441


RESULT 12
US-09-152-060-19
; Sequence 19, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 28 Human Secreted Proteins
;  FILE REFERENCE: PZ003P1.US
;  CURRENT APPLICATION NUMBER: US/09/152,060
;  CURRENT FILING DATE: 1998-09-11
;  EARLIER APPLICATION NUMBER: PCT/US98/04858
;  EARLIER FILING DATE: 1998-03-12
;  EARLIER APPLICATION NUMBER: 60/040,762
;  EARLIER FILING DATE: 1997-03-14
;  EARLIER APPLICATION NUMBER: 60/040,710
;  EARLIER FILING DATE: 1997-03-14
;  EARLIER APPLICATION NUMBER: 60/050,934
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/048,100
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/048,357
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/048,189
;  EARLIER FILING DATE: 1997-05-30
;  EARLIER APPLICATION NUMBER: 60/057,765
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/048,970
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/068,368
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 118
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 1699
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (871)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-152-060-19

  Query Match             2.0%;  Score 56.4;  DB 4;  Length 1699;
  Best Local Similarity   80.5%;  Pred. No. 9.8e-06;
  Matches   66;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    2694 ATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAAAAA 2753
           |||||| |     ||| |||  ||||  |   || || ||||||||||||||||||||||
Db    1610 ATGGCTTTCCGTAATAAAAAGATTGGGATTTCCTTTTGAAAAAAAAAAAAAAAAAAAAAAA 1669

Qy    2754 AAAAAAAAAAAAAAAAAAAAA 2775
           |||||||||||||||||||||
Db    1670 AAAAAAAAAAAAAAAAAAAAA 1691


RESULT 13
US-09-196-390-1
; Sequence 1, Application US/09196390
; Patent No. 6307125
; GENERAL INFORMATION:
;    APPLICANT:  Block, Martina
;    APPLICANT:  Lorz, Horst
;    APPLICANT:  Lutticke, Stephanie
;    APPLICANT:  Walter, Lennart
;    APPLICANT:  Frohberg, Claus
;    APPLICANT:  Kossmann, Jens
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING ENZYMES
;    TITLE OF INVENTION:  FROM WHEAT WHICH ARE INVOLVED IN STARCH
;    TITLE OF INVENTION:  SYNTHESIS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James F. Haley, Jr., c/o Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/196,390
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 196 21 588.9
;      FILING DATE:  29-MAY-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 196 36 917.7
;      FILING DATE:  11-SEP-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/02793
;      FILING DATE:  28-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, Jr., James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  AGREVO-9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 596-9000
;      TELEFAX:  (212) 596-9090
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2239 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Triticum aestivum L.
;      STRAIN:  cv. Florida
;      HAPLOTYPE:  ca. 21 d Caryopses
;    IMMEDIATE SOURCE:
;      LIBRARY:  cDNA library in pBluescript sk (-)
;      CLONE:  TaSSS
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..2017
US-09-196-390-1

  Query Match             2.0%;  Score 56.4;  DB 4;  Length 2239;
  Best Local Similarity   63.0%;  Pred. No. 1.2e-05;
  Matches   87;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy   2638 TCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCTCACATGG 2697
          | ||  |   |||||  ||  | |   ||  ||| || ||| |  || ||  |  |   |
Db   2088 TGTTGGAGAATTTTATCTTTGCTGCTGTTTTTTTTTAATCAAAAGAGGGGGTTTCCTCCG 2147

Qy   2698 CTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAA 2757
           | |     | | |||      |  |||     ||||||||||||||||||||||||||
Db   2148 ATTTCATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2207

Qy   2758 AAAAAAAAAAAAAAAAA 2775
          |||||||||||||||||
Db   2208 AAAAAAAAAAAAAAAAA 2225


RESULT 14
US-09-149-476-313
; Sequence 313, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 186 Human Secreted proteins
;  FILE REFERENCE: PZ002P1
;  CURRENT APPLICATION NUMBER: US/09/149,476
;  CURRENT FILING DATE: 1998-09-08
;  EARLIER APPLICATION NUMBER: PCT/US98/04493
;  EARLIER FILING DATE: 1998-03-06
;  EARLIER APPLICATION NUMBER: 60/040,162
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,333
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/038,621
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,626
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,334
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,336
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/040,163
;  EARLIER FILING DATE: 1997-03-07
;  EARLIER APPLICATION NUMBER: 60/047,600
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,615
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,597
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,502
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,633
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,583
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,617
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,618
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,503
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,592
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,581
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,584
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,500
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,587
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,492
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,598
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,613
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,582
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,596
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,612
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,632
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,601
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/043,580
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,568
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,314
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,569
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,311
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,671
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,674
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,669
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,312
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,313
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,672
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,315
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/048,974
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/056,886
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,877
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,889
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,893
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,630
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,878
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,662
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,872
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,882
;  EARLIER FILING DATE: 1997-08-22
```

```
;  EARLIER APPLICATION NUMBER: 60/056,637
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,903
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,888
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,879
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,880
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,894
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,911
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,636
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,874
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,910
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,864
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,631
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,845
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,892
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/057,761
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/047,595
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,599
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,588
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,585
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,586
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,590
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,594
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,589
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,593
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,614
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/043,578
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,576
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/047,501
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/043,670
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/056,632
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,664
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,876
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,881
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,909
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,875
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,862
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,887
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,908
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/057,650
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/056,884
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/057,669
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/049,610
;  EARLIER FILING DATE: 1997-06-13
;  EARLIER APPLICATION NUMBER: 60/061,060
;  EARLIER FILING DATE: 1997-10-02

  Query Match             2.0%;  Score 56.2;  DB 4;  Length 577;
  Best Local Similarity   59.5%;  Pred. No. 5.7e-06;
  Matches   91;  Conservative    2;  Mismatches   60;  Indels    0;  Gaps    0;

Qy    2623 TATAGAAATGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGA 2682
           ||  | |  |||||   ||||||||  |   | |       |  |  |  | |   | |
Db     415 TACTGGAGGGATTTCATTTTTAATTCATCTTTATGAAGATTTAGAACTCATTCCTTGTGT 474

Qy    2683 AGAGGCCTCACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAA 2742
             |      |  |      ||  | ||| | || |  |:  ||| | :||||||||||||
Db     475 TTAAAGGGAATGTTTAATTGAGAAATAAACATTTGTGWACAAAATGYTAAAAAAAAAAAA 534

Qy    2743 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2775
           ||||||||||||||||||||||||||||||||
Db     535 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 567


RESULT 15
US-08-464-517-1
; Sequence 1, Application US/08464517
; Patent No. 5869640
;  GENERAL INFORMATION:
;    APPLICANT:  BEACH, David H.
;    TITLE OF INVENTION:  D-TYPE CYCLIN AND USES RELATED THERETO
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII(text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,517
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/963,308
;      FILING DATE:  16-OCT-1992
;      APPLICATION NUMBER:  US 07/888,178
;      FILING DATE:  26-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/701,514
;      FILING DATE:  16-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Matthew P. Vincent
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  MII-004C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  145..1029
US-08-464-517-1

  Query Match              2.0%;  Score 55.4;  DB 2;  Length 1325;
  Best Local Similarity   61.4%;  Pred. No. 1.6e-05;
  Matches   89;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy    2631 TGATTTTTCTTTTAATTTTTTACTTTACCAGACTTTACTTTGTACTCAGAGAAGAGGCCT 2690
           || |||||| |||  | |||  |  | |||    | ||||       | | ||
Db    1167 TGTTTTTTCCTTTGCTCTTTCCCCCTTCCATCTCTCACTTAACCAAAACAAAAAGATTAC 1226

Qy    2691 CACATGGCTGTGTCACATATAAATGTTGGACTAAACTCTTAAAAAAAAAAAAAAAAAAAA 2750
           |  |   |||| |   | | | |    |    |||     ||||||||||||||||||||
Db    1227 CCAAAAACTGTCTTTAAAAGAGAGAGAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1286

Qy    2751 AAAAAAAAAAAAAAAAAAAAAAAAA 2775
           |||||||||||||||||||||||||
Db    1287 AAAAAAAAAAAAAAAAAAAAAAAAA 1311


Search completed: April 11, 2003, 22:41:26
Job time : 111.627 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 21:30:39 ; Search time 177.543 Seconds
(without alignments)
13710.183 Million cell updates/sec

Title: US-09-914-152-2
Perfect score: 2775
Sequence: 1 gtgaattcctctttctctgc.....aaaaaaaaaaaaaaaaaaa 2775

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	130.4	4.7	1773	10	US-09-739-451-9	Sequence 9, Appli
2	130.4	4.7	2095	9	US-09-992-598-208	Sequence 208, App
3	130.4	4.7	2095	9	US-09-989-293A-208	Sequence 208, App
4	130.4	4.7	2095	9	US-10-063-547-35	Sequence 35, Appl
5	130.4	4.7	2095	9	US-09-989-735-208	Sequence 208, App
6	130.4	4.7	2095	9	US-09-990-444-208	Sequence 208, App
7	130.4	4.7	2095	9	US-09-989-730-208	Sequence 208, App
8	130.4	4.7	2095	9	US-09-990-436-208	Sequence 208, App
9	130.4	4.7	2095	9	US-09-991-181-208	Sequence 208, App
10	130.4	4.7	2095	9	US-09-993-687-208	Sequence 208, App
11	130.4	4.7	2095	9	US-09-989-734-208	Sequence 208, App
12	130.4	4.7	2095	9	US-09-997-653-208	Sequence 208, App
13	130.4	4.7	2095	9	US-10-174-590-161	Sequence 161, App
14	130.4	4.7	2095	9	US-10-176-758-161	Sequence 161, App
15	130.4	4.7	2095	9	US-10-063-616-35	Sequence 35, Appl
16	130.4	4.7	2095	9	US-10-175-737-161	Sequence 161, App
17	130.4	4.7	2095	9	US-09-993-667-208	Sequence 208, App
18	130.4	4.7	2095	9	US-10-063-502-35	Sequence 35, Appl
19	130.4	4.7	2095	9	US-10-173-706-161	Sequence 161, App
20	130.4	4.7	2095	9	US-10-175-738-161	Sequence 161, App
21	130.4	4.7	2095	9	US-10-175-752-161	Sequence 161, App
22	130.4	4.7	2095	9	US-10-176-482-161	Sequence 161, App
23	130.4	4.7	2095	9	US-10-176-757-161	Sequence 161, App
24	130.4	4.7	2095	9	US-10-176-913-161	Sequence 161, App
25	130.4	4.7	2095	9	US-10-180-552-161	Sequence 161, App
26	130.4	4.7	2095	9	US-10-180-557-161	Sequence 161, App
27	130.4	4.7	2095	9	US-09-990-438-208	Sequence 208, App
28	130.4	4.7	2095	9	US-09-990-562-208	Sequence 208, App
29	130.4	4.7	2095	9	US-09-997-428-208	Sequence 208, App
30	130.4	4.7	2095	9	US-09-997-666-208	Sequence 208, App
31	130.4	4.7	2095	9	US-10-173-700-161	Sequence 161, App
32	130.4	4.7	2095	9	US-10-174-572-161	Sequence 161, App
33	130.4	4.7	2095	9	US-10-174-579-161	Sequence 161, App
34	130.4	4.7	2095	9	US-10-174-582-161	Sequence 161, App
35	130.4	4.7	2095	9	US-10-174-588-161	Sequence 161, App
36	130.4	4.7	2095	9	US-10-175-739-161	Sequence 161, App
37	130.4	4.7	2095	9	US-10-175-740-161	Sequence 161, App
38	130.4	4.7	2095	9	US-10-175-743-161	Sequence 161, App
39	130.4	4.7	2095	9	US-10-176-488-161	Sequence 161, App
40	130.4	4.7	2095	9	US-10-176-492-161	Sequence 161, App
41	130.4	4.7	2095	9	US-10-176-747-161	Sequence 161, App
42	130.4	4.7	2095	9	US-10-176-750-161	Sequence 161, App
43	130.4	4.7	2095	9	US-10-176-985-161	Sequence 161, App
44	130.4	4.7	2095	9	US-10-176-987-161	Sequence 161, App
45	130.4	4.7	2095	9	US-10-176-991-161	Sequence 161, App

ALIGNMENTS

RESULT 1
US-09-739-451-9
; Sequence 9, Application US/09739451
; Patent No. US20010024813A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Dendriac and Brainiac-3
; FILE REFERENCE: PF464
; CURRENT APPLICATION NUMBER: US/09/739,451
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/213,364
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/108,928
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1773
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-739-451-9

Query Match 4.7%; Score 130.4; DB 10; Length 1773;
Best Local Similarity 48.9%; Pred. No. 2.5e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||  |||||  |||||       |       || || |||||  ||   |  ||
Db     157 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 216

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | ||||||||||||   ||    | |  ||   |   |  ||||||
Db     217 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 276

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || ||||| || | | |     |  |||| |  || |    |||||  ||| |
Db     277 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 336

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                | || | |   ||         | ||  | ||   ||
```

```
Db    337 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 396

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801

Db    397 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 456

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861

Db    457 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 516

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921

Db    517 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 576

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981

Db    577 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 636

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041

Db    637 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 696

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101

Db    697 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 756

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161

Db    757 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 816

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221

Db    817 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 876

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281

Db    877 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 936

Qy   1282 ACTACTGGCAGG 1293

Db    937 CTTTTTGGCAGG 948


RESULT 2
US-09-992-598-208
; Sequence 208, Application US/09992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C20
;  CURRENT APPLICATION NUMBER: US/09/992,598
;  CURRENT FILING DATE:  2001-11-14
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088810
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088858
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088861
```

; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25

; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;
Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||     | |  ||    |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |      |  |||| |  || |    |||| ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                |  || |  |    ||         | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
          |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |     |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |        | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
```

```
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
             || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
             |  |||||||
Db    1274 CTTTTTGGCAGG 1285


RESULT 3
US-09-989-293A-208
; Sequence 208, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C66
;  CURRENT APPLICATION NUMBER: US/09/989,293A
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088810
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088858
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088861
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088876
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089440
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089512
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089532
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089538
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089598
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089599
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089600
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089653
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089801
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/089907
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/089908
```

; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;

Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||    ||     | |  ||    |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |      |  |||| |  || |    |||||  ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                |  || |  |    ||         | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
            |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   |||||  ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
            |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
            |  |||||||
Db   1274 CTTTTTGGCAGG 1285
```

RESULT 4
US-10-063-547-35
; Sequence 35, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
; APPLICANT: Eaton,Dan L.
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K.

```
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,547
;  CURRENT FILING DATE:  2002-05-02
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 35
;   LENGTH: 2095
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-063-547-35

  Query Match             4.7%;  Score 130.4;  DB 9;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 2.8e-25;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||       |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |     |  |||| |  || |    |||||  ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                |  || |  |    ||          | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
            |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |     |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||| || |       | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
          | |   ||     | |  | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||   |  | ||  | ||  ||| |     | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
            |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  |  | ||   |||   | |      |   |    || |
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
            |  |||||||
Db   1274 CTTTTTGGCAGG 1285
```

```
RESULT 5
US-09-989-735-208
; Sequence 208, Application US/09989735
; Publication No. US20020193299A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C61
;  CURRENT APPLICATION NUMBER: US/09/989,735
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
```

; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;
Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||       |       || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |     |  |||| |  || |    ||||| ||| |
```

```
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741

Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801

Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861

Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921

Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981

Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041

Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101

Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161

Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221

Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281

Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293

Db   1274 CTTTTTGGCAGG 1285
```

RESULT 6
US-09-990-444-208
; Sequence 208, Application US/09990444
; Publication No. US20020193300A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi,Avi J.
; APPLICANT: Baker,Kevin P.
; APPLICANT: Botstein,David
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Fong,Sherman
; APPLICANT: Gerber,Hanspeter
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,J.Christopher
; APPLICANT: Gurney,Austin L.
; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C19
; CURRENT APPLICATION NUMBER: US/09/990,444
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826

; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;
Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |      |  |||| |  || |    ||||| ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||               |  || |  |    ||          | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
            |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | ||||||||| || | || |      |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||| || |      | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
```

```
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161

Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221

Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281

Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293

Db    1274 CTTTTTGGCAGG 1285


RESULT 7
US-09-989-730-208
; Sequence 208, Application US/09989730
; Publication No. US20020197674A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C69
;  CURRENT APPLICATION NUMBER: US/09/989,730
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088810
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088858
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088861
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088876
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089440
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089512
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089532
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089538
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089598
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089599
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089600
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089653
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089801
```

; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;
Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||       |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | | ||   |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |     |  |||| |  || |    ||||| ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||              |  || |  |    ||         | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
            |  | || ||||| ||||  | |||||  |||||  ||| | || |||||| |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |     |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||      | |   | |    || ||||| |    |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |      | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
            |||   |  ||| ||  | ||||| || | ||| ||| |||| |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |  |   |          || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
            |  |||||||
Db   1274 CTTTTTGGCAGG 1285
```

RESULT 8
US-09-990-436-208
; Sequence 208, Application US/09990436
; Publication No. US20020198148A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi,Avi J.
; APPLICANT: Baker,Kevin P.
; APPLICANT: Botstein,David
; APPLICANT: Desnoyers,Luc

; APPLICANT: Eaton,Dan L.
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Fong,Sherman
; APPLICANT: Gerber,Hanspeter
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,J.Christopher
; APPLICANT: Gurney,Austin L.
; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C14
; CURRENT APPLICATION NUMBER: US/09/990,436
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444

```
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090445
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090472
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090535
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090540
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090542
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090557
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090676
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090678
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090690
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090694
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090695
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090696
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090862
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/090863
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/091360
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091478
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091544
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091519
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091626
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091633
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091978
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/091982
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/092182
;  PRIOR FILING DATE: 1998-07-09

  Query Match             4.7%;  Score 130.4;  DB 9;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 2.8e-25;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | | |     |  |||| |  || |    |||||  ||| |
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                |  || |  |    ||         | ||  | ||   ||
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
             |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           |  ||||| |||||    |  |||||  |||||  |||||| ||  | ||  || || ||||
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
             | |   ||       | |   | |    || ||||| |   |   || ||| |||| ||
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| |||| |  |    | |
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
             || || |   | ||  |  |  | ||  |||   | |     |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
             |  |||||||
Db    1274 CTTTTGGCAGG 1285


RESULT 9
US-09-991-181-208
; Sequence 208, Application US/09991181
; Publication No. US20020197615A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C53
;  CURRENT APPLICATION NUMBER: US/09/991,181
;  CURRENT FILING DATE:  2001-11-16
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
```

; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863

```
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/091360
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091478
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091544
;  PRIOR FILING DATE: 1998-07-01
;  PRIOR APPLICATION NUMBER: 60/091519
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091626
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091633
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/091978
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/091982
;  PRIOR FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: 60/092182
;  PRIOR FILING DATE: 1998-07-09

  Query Match             4.7%;  Score 130.4;  DB 9;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 2.8e-25;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||       |       || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||    ||     | |   ||    |    |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |      |  |||| |  || |    |||||  ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                |  || |  |    ||         | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
            |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |      |   ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
            |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |  | |    |   |   |          || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
```

```
          || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
            |  |||||||
Db   1274 CTTTTTGGCAGG 1285


RESULT 10
US-09-993-687-208
; Sequence 208, Application US/09993687
; Publication No. US20020198149A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C11
;  CURRENT APPLICATION NUMBER: US/09/993,687
;  CURRENT FILING DATE:  2002-11-14
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
```

; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;
Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

Qy 511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
|||||| ||||| ||||| | || || ||||| || | ||
Db 494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

```
Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | ||||||||||||   ||     | |  ||   |   |   ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | | |      |  |||| |  || |    |||||  ||| |
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||              |  || |  |    ||          | ||  | ||   ||
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
             |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |     |    ||| | ||   ||
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
             | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |  |  |          || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
             || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
             |  |||||||
Db    1274 CTTTTTGGCAGG 1285


RESULT 11
US-09-989-734-208
; Sequence 208, Application US/09989734
; Publication No. US20030003531A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C64
;  CURRENT APPLICATION NUMBER: US/09/989,734
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
```

; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;
Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | |  ||   |   |  ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |     |  |||| |  || |    |||||  ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                |  || |  |   ||         | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
          |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |   |||||   |  ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973
```

```
Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
           | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
           || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
           |  |||||||
Db    1274 CTTTTTGGCAGG 1285


RESULT 12
US-09-997-653-208
; Sequence 208, Application US/09997653
; Publication No. US20030008297A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi,Avi J.
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Botstein,David
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Ferrara,Napoleone
;  APPLICANT:  Fong,Sherman
;  APPLICANT:  Gerber,Hanspeter
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,J.Christopher
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Kljavin,Ivar J.
;  APPLICANT:  Napier,Mary A.
;  APPLICANT:  Pan,James
;  APPLICANT:  Paoni,Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Watanabe, Colin K.
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2730P1C38
;  CURRENT APPLICATION NUMBER: US/09/997,653
;  CURRENT FILING DATE:  2001-11-15
;  PRIOR APPLICATION NUMBER: 60/049787
;  PRIOR FILING DATE: 1997-06-16
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/065186
;  PRIOR FILING DATE: 1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE: 1997-11-24
;  PRIOR APPLICATION NUMBER: 60/075945
;  PRIOR FILING DATE: 1998-02-25
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/083322
;  PRIOR FILING DATE: 1998-04-28
;  PRIOR APPLICATION NUMBER: 60/084600
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 60/087106
;  PRIOR FILING DATE: 1998-05-28
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087609
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/087827
;  PRIOR FILING DATE: 1998-06-03
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088025
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088026
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088028
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088033
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088167
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088202
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088742
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088810
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088858
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088861
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088876
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089440
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089512
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089532
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089538
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089598
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089599
```

; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.7%; Score 130.4; DB 9; Length 2095;
Best Local Similarity 48.9%; Pred. No. 2.8e-25;
Matches 387; Conservative 0; Mismatches 396; Indels 9; Gaps 1;

```
Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||       |       || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |     |  |||| |  || |    ||||| ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                 |  || |  |    ||         | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
            |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |     |   ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||       | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||    ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
            |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |  |          || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  |  | ||   |||   | |     |   |  | | ||
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
            |  |||||||
Db   1274 CTTTTTGGCAGG 1285
```

RESULT 13
US-10-174-590-161
; Sequence 161, Application US/10174590

```
; Publication No. US20030008352A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C42
;  CURRENT APPLICATION NUMBER: US/10/174,590
;  CURRENT FILING DATE:  2002-06-18
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 161
;   LENGTH: 2095
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-174-590-161

  Query Match             4.7%;  Score 130.4;  DB 9;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 2.8e-25;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||   |||||    |||||       |       || || |||||   ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | ||||||||||   ||     | |  ||   |   |   ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | |      |  |||| |  || |    |||||  ||| |
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                 |  || |  |    ||         | ||  | ||   ||
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
           |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           | ||||| ||||   |    |||||    |  ||||| ||  | ||   ||  || ||||
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
           ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
           |  |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
              |||   |  ||| ||  | ||||| || | ||| ||| |||| |  |     | |
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
           |  | |   |  |    |   |      || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
             || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    1282 ACTACTGGCAGG 1293
             |  |||||||
Db    1274 CTTTTTGGCAGG 1285


RESULT 14
US-10-176-758-161
; Sequence 161, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C104
;  CURRENT APPLICATION NUMBER: US/10/176,758
;  CURRENT FILING DATE:  2002-06-21
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 161
;   LENGTH: 2095
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-176-758-161

  Query Match             4.7%;  Score 130.4;  DB 9;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 2.8e-25;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
           ||||||   |||||    |||||       |       || || |||||   ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
           | || || ||| | ||||||||||   ||     | |  ||   |   |   ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
           |    || |||||  || | |      |  |||| |  || |    |||||  ||| |
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
           | ||                 |  || |  |    ||         | ||  | ||   ||
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
           |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
           | |||||  ||||    |    |||||    |  ||||| ||  | ||   ||  || ||||
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
           | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
```

```
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |      | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
              |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  | || |||   | |      |   |    || |
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
             |  |||||||
Db   1274 CTTTTTGGCAGG 1285


RESULT 15
US-10-063-616-35
; Sequence 35, Application US/10063616
; Publication No. US20030013855A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,616
;  CURRENT FILING DATE:  2002-05-03
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 35
;   LENGTH: 2095
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-063-616-35

  Query Match             4.7%;  Score 130.4;  DB 9;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 2.8e-25;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    511 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 570
          ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    571 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 630
          | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    631 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 690
          |    || |||||  || | | |     |  |||| |  || |    ||||| ||| |
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    691 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 741
          | ||                  |  || |  |    ||         | ||  | ||   ||
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    742 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 801
          |   | || ||||| |||| | |||||  |||||   ||| | || |||||| |||
Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    802 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 861
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    862 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 921
          | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    922 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 981
          ||    ||   | |||||||||| || |      | |  || | | | ||   ||
Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    982 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 1041
            | |   ||      | |   | |    || ||||| |   |   || ||| |||| ||
Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy   1042 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 1101
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy   1102 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 1161
              |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy   1162 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 1221
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy   1222 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 1281
            || || |   | ||  |  | || |||   | |      |   |    || |
Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy   1282 ACTACTGGCAGG 1293
             |  |||||||
Db   1274 CTTTTTGGCAGG 1285


Search completed: April 12, 2003, 03:33:18
Job time : 184.543 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2003, 12:08:21 ; Search time 33 Seconds
(without alignments)
1935.594 Million cell updates/sec

Title: US-09-914-152-1
Perfect score: 1657
Sequence: 1 MAFPKMRLMYICLLVLGALC.....TLLDYWQALENSRGEDCPPV 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1541.5	93.0	302	6	Q9N292	Q9n292 pongo pygma
2	589	35.5	326	11	O54904	O54904 mus musculu
3	587	35.4	326	4	Q9Y5Z6	Q9y5z6 homo sapien
4	587	35.4	326	6	Q9MYM7	Q9mym7 pongo pygma
5	576	34.8	305	11	Q920V5	Q920v5 mus spicile
6	576	34.8	305	11	Q91V52	Q91v52 mus musculu
7	569	34.3	422	4	O43825	O43825 homo sapien
8	567	34.2	409	11	Q920V4	Q920v4 mus musculu
9	567	34.2	409	11	Q920V3	Q920v3 mus musculu
10	567	34.2	409	11	Q920V2	Q920v2 mus spicile
11	567	34.2	409	11	Q91VE9	Q91ve9 mus musculu
12	567	34.2	409	11	Q91V58	Q91v58 mus musculu
13	567	34.2	409	11	Q91V19	Q91v19 mus musculu
14	567	34.2	422	11	O54905	O54905 mus musculu
15	565	34.1	422	4	Q9BZQ9	Q9bzq9 homo sapien
16	558	33.7	331	11	Q920V1	Q920v1 mus musculu
17	558	33.7	331	11	Q920V0	Q920v0 mus musculu
18	558	33.7	331	11	Q91V72	Q91v72 mus musculu
19	558	33.7	331	11	O54906	O54906 mus spicile
20	555.5	33.5	331	4	O75752	O75752 homo sapien
21	529.5	32.0	255	11	Q9CTE5	Q9cte5 mus musculu
22	434.5	26.2	412	13	Q8UWM2	Q8uwm2 brachydanio
23	429	25.9	204	4	Q8TDY1	Q8tdy1 homo sapien
24	416	25.1	372	4	Q9C0J2	Q9c0j2 homo sapien
25	407.5	24.6	390	13	Q8UWM1	Q8uwm1 brachydanio
26	403.5	24.4	372	11	Q9D7Z2	Q9d7z2 mus musculu
27	402.5	24.3	372	11	Q8R0U2	Q8r0u2 mus musculu
28	396	23.9	378	4	Q9BYG0	Q9byg0 homo sapien
29	393	23.7	371	11	O88178	O88178 rattus norv
30	388	23.4	370	11	Q920U7	Q920u7 mus spicile
31	388	23.4	370	11	Q91VC1	Q91vc1 mus musculu
32	388	23.4	371	11	Q9Z0F0	Q9z0f0 mus musculu
33	388	23.4	378	4	O96024	O96024 homo sapien
34	388	23.4	406	13	Q8UWM0	Q8uwm0 brachydanio
35	386	23.3	418	13	Q8UWM4	Q8uwm4 brachydanio
36	383.5	23.1	384	4	Q8TDX1	Q8tdx1 homo sapien
37	383	23.1	370	11	Q920U8	Q920u8 mus musculu
38	382.5	23.1	277	4	Q8TAZ4	Q8taz4 homo sapien
39	381.5	23.0	574	5	Q9VLS8	Q9vls8 drosophila
40	381.5	23.0	585	5	Q95TQ8	Q95tq8 drosophila
41	380	22.9	370	11	Q920U9	Q920u9 mus musculu
42	364	22.0	412	13	Q8UWM3	Q8uwm3 brachydanio
43	362	21.8	399	11	Q8VI16	Q8vi16 mus musculu
44	359.5	21.7	377	4	Q96EK0	Q96ek0 homo sapien
45	359	21.7	397	11	Q91V18	Q91v18 mus musculu

ALIGNMENTS

RESULT 1
Q9N292
ID Q9N292 PRELIMINARY; PRT; 302 AA.
AC Q9N292;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE UDP-Gal:GlcNAc beta1,3-galactosyltransferase 5 (Fragment).
GN BETA1,3-GALT 5.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORAN-PO17;
RA Liu Y., Saitou N.;
RT "Silver Project.";
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB041417; BAA94502.1; -.
DR InterPro; IPR002659; Galactosyl_T.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Glycosyltransferase; Transferase.
FT NON_TER 302 302
SQ SEQUENCE 302 AA; 35500 MW; 51A555C864465A90 CRC64;

Query Match 93.0%; Score 1541.5; DB 6; Length 302;
Best Local Similarity 96.4%; Pred. No. 1.8e-137;
Matches 291; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

```
Qy       1 MAFPKMRLMYICLLVLGALCLYFSMYSLNPFKEQSFVYKK-DGNFLKLPDTDCRQTPPFL 59
           |||||||:|||||||||||||||:|||||||||||| ||||||||||||||||||||
Db       1 MAFPKMRLIYICLLVLGALCLYFSMYNLNPFKEQSFVYKKEDGNFLKLPDTDCRQTPPFL 60

Qy      60 VLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDI 119
           |||||||||||||||||||||||| |||||||||||||||||| ||||||||||||
Db      61 VLLVTSSHKQLAERMAIRQTWGKERTVKGKQLKTFFLLGTTSSAAEMKEVDQESQRHGDI 120

Qy     120 IQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRF 179
```

```
Db    121 IQKDFLDVYYNLTLKTMMGMEWVHRFCPQAAFVMKTDSDMFINVDYLTELLLKKNRTTRF 180

Qy    180 FTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYI 239

Db    181 FTGFLKLNELPIRQPFSKWFVSKSEYPWDRYPPFCSGTAYVFSGDVASQVYNVSESVPYI 240

Qy    240 KLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFIKPRTLLDYWQAL 299

Db    241 KLEDVFVGLCLERLNIRLEELHSQPTFFPEGLRFSVCRFRRIVACHFIKPRTLLDYWQAL 300

Qy    300 EN 301

Db    301 EN 302
```

RESULT 2
O54904
ID O54904 PRELIMINARY; PRT; 326 AA.
AC O54904;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE II (EC 2.4.1.-) (UDP-GAL:BETAGLCNAC BETA 1,3-GALACTOSYLTRANFERASE-I)
DE (BETA3GALT-I).
GN B3GT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=129/SVJ;
RX MEDLINE=98079027; PubMed=9417047;
RA Hennet T., Dinter A., Kuhnert P., Mattu T.S., Rudd P.M., Berger E.G.;
RT "Genomic cloning and expression of three murine UDP-galactose: beta-N-
RT acetylglucosamine beta1,3-galactosyltransferase genes.";
RL J. Biol. Chem. 273:58-65(1998).
CC -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT-I, GALT-II
CC AND GALT-III ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH
CC DIFFERENT AFFINITY.
CC -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC GLUCOSAMINYLGLYCOPEPTIDE.
CC -!- COFACTOR: MANGANESE.
CC -!- PATHWAY: GLYCOSYLATION.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND TO A LESSER EXTENT IN
CC HEART, KIDNEY, LIVER, LUNG, OVARY, TESTIS, UTERUS, COLON, STOMACH,
CC BONE MARROW, THYMUS, SPLEEN AND LYMPH NODES.
DR EMBL; AF029790; AAC53523.1; -.
DR MGD; MGI:1349403; B3galt1.
DR InterPro; IPR002659; Galactosyl_T.
DR Pfam; PF01762; Galactosyl_T; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Glycoprotein; Golgi stack.
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 26 326 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 326 AA; 37979 MW; 9825D68869EFAD3D CRC64;

Query Match 35.5%; Score 589; DB 11; Length 326;
Best Local Similarity 38.6%; Pred. No. 1.9e-47;
Matches 127; Conservative 52; Mismatches 108; Indels 42; Gaps 7;

```
Qy     12 CLLVLGALCLYFSMYSLN-----------PFKEQSFVYKK--DGN-----------FL 45

Db      7 CLYVLSVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFL 66

Qy     46 KLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE 105

Db     67 INEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVL 126

Qy    106 TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDY 165

Db    127 NQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDN 186

Qy    166 LTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSG 223

Db    187 LIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSA 245

Qy    224 DVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVCLF 278

Db    246 DVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLCRY 297

Qy    279 RRIVACHFIKPRTLLDYWQALENSRGEDC 307

Db    298 RRVITVHQISPEEMHRIWNDMSSKKHLRC 326
```

RESULT 3
Q9Y5Z6
ID Q9Y5Z6 PRELIMINARY; PRT; 326 AA.
AC Q9Y5Z6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE 1 (EC 2.4.1.-) (BETA3GALT1).
GN B3GALT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=COLON ADENOCARCINOMA;
RX MEDLINE=99283878; PubMed=10356986;
RA Bardoni A., Valli M., Trinchera M.;
RT "Differential expression of beta1,3galactosyltransferases in human
RT colon cells derived from adenocarcinomas or normal mucosa.";
RL FEBS Lett. 451:75-80(1999).
RN [2]
RP FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=98250717; PubMed=9582303;
RA Amado M., Almeida R., Carneiro F., Levery S.B., Holmes E.H.,
RA Nomoto M., Hollingsworth M.A., Hassan H., Schwientek T., Nielsen P.A.,
RA Bennett E.P., Clausen H.;
RT "A family of human beta3-galactosyltransferases. Characterization of
RT four members of a UDP-galactose:beta-N-acetyl-glucosamine/beta-
RT nacetyl-galactosamine beta-1,3-galactosyltransferase family.";
RL J. Biol. Chem. 273:12770-12778(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Liu Y., Saitou N.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT1 AND GALT2
CC ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH DIFFERENT
CC AFFINITY.
CC -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC GLUCOSAMINYLGLYCOPEPTIDE.
CC -!- PATHWAY: GLYCOSYLATION.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND COLON MUCOSA AND TO A
CC LESSER EXTENT IN COLON ADENOCARCINOMA CELLS. NO EXPRESSION IN
CC HEART, PLACENTA, KIDNEY, LIVER, SKELETAL MUSCLE, LUNG AND
CC PANCREAS.
DR EMBL; AF117222; AAD23451.1; -.

```
DR   EMBL; AB041407; BAA94492.1; -.
DR   MIM; 603093; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Glycoprotein; Golgi stack.
FT   DOMAIN        1      4       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      5     20       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       21    326       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   326 AA;  37993 MW;  83271E99B2EE74F5 CRC64;

  Query Match             35.4%;  Score 587;  DB 4;  Length 326;
  Best Local Similarity   38.6%;  Pred. No. 3e-47;
  Matches  127;  Conservative   52;  Mismatches  108;  Indels   42;  Gaps    7;

Qy      12 CLLVLGALCLYFSMYSLN-----------PFKEQSFVYKK--DGN-----------FL 45
           || ||  :|   ::: |:            ||   :   |    ||           ||
Db       7 CLYVLTVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFL 66

Qy      46 KLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE 105
              | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db      67 INEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVL 126

Qy     106 TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDY 165
           : |:||||   ||| :||:| |:||||||:|| ||  || :| :|||||||:|:|:|
Db     127 NQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDN 186

Qy     166 LTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSG 223
           |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:||
Db     187 LIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSA 245

Qy     224 DVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVCLF 278
           |||  :|  |     : ||||:||||| :| |          |   |     : :|:| :
Db     246 DVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLCRY 297

Qy     279 RRIVACHFIKPRTLLDYWQALENSRGEDC 307
           ||::  | | |  :   |  : : :   |
Db     298 RRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 4
Q9MYM7
ID   Q9MYM7      PRELIMINARY;      PRT;   326 AA.
AC   Q9MYM7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Beta 1,3-galactosyltransferase polypeptide 1.
GN   B3GALT1.
OS   Pongo pygmaeus (Orangutan),
OS   Pan troglodytes (Chimpanzee),
OS   Pan paniscus (Pygmy chimpanzee) (Bonobo), and
OS   Gorilla gorilla (gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600, 9598, 9597, 9593;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=P.pygmaeus, P.troglodytes, P.paniscus, and G.gorilla;
RA   Liu Y., Saitou N.;
RT   "Silver Project.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB041411; BAA94496.1; -.
DR   EMBL; AB041408; BAA94493.1; -.
DR   EMBL; AB041409; BAA94494.1; -.
DR   EMBL; AB041410; BAA94495.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   326 AA;  37993 MW;  83271E99B2EE74F5 CRC64;

  Query Match             35.4%;  Score 587;  DB 6;  Length 326;
  Best Local Similarity   38.6%;  Pred. No. 3e-47;
  Matches  127;  Conservative   52;  Mismatches  108;  Indels   42;  Gaps    7;

Qy      12 CLLVLGALCLYFSMYSLN-----------PFKEQSFVYKK--DGN-----------FL 45
           || ||  :|   ::: |:            ||   :   |    ||           ||
Db       7 CLYVLTVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFL 66

Qy      46 KLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTTSSAAE 105
              | :  ||||:|::::||:   | |||:||| |   || :: | ||||  :
Db      67 INEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGKNADPVL 126

Qy     106 TKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDY 165
           : |:||||   ||| :||:| |:||||||:|| ||  || :| :|||||||:|:|:|
Db     127 NQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDIFVNMDN 186

Qy     166 LTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSG 223
           |   ||| :     |:|||:: :|  |||   |||:: :  ||   ||||||||||:||
Db     187 LIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSA 245

Qy     224 DVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGG-----LRFSVCLF 278
           |||  :|  |     : ||||:||||| :| |          |   |     : :|:| :
Db     246 DVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH--------PFQNSGFNHWKMAYSLCRY 297

Qy     279 RRIVACHFIKPRTLLDYWQALENSRGEDC 307
           ||::  | | |  :   |  : : :   |
Db     298 RRVITVHQISPEEMHRIWNDMSSKKHLRC 326


RESULT 5
Q920V5
ID   Q920V5      PRELIMINARY;      PRT;   305 AA.
AC   Q920V5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase I (Fragment).
GN   B3GT1.
OS   Mus spicilegus (Steppe mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZBN;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039143; BAB68667.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     305    305
SQ   SEQUENCE   305 AA;  35611 MW;  A338019F3B06EF5A CRC64;

  Query Match             34.8%;  Score 576;  DB 11;  Length 305;
  Best Local Similarity   42.3%;  Pred. No. 3e-46;
  Matches  119;  Conservative   43;  Mismatches   97;  Indels   22;  Gaps    5;

Qy      23 FSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGK 82
           |      |    ||       ||      | :  ||||:|::::||:   | |||:|||
Db      39 FGNIRTRPINPHSF------EFLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGD 92

Qy      83 ERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWV 142
           |   || :: | ||||  :     : |:||||   ||| :||:| |:||||||:||: ||
Db      93 ENNFKGIKIATLFLLGKNADPVLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWV 152

Qy     143 HRFCPQAAFVMKTDSDMFINVDYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFV 200
             || :| :|||||||:|:|:| |   ||| :     |:|||:: :|  |||   |||::
Db     153 ATFCSKAKYVMKTDSDIFVNMDNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYM 211
```

```
Qy     201 SKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEEL 260
           :  ||   ||||||||||:|| |||  :|  |     : ||||:||||| :| |
Db     212 PRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH---- 267

Qy     261 HSQPTFFPGG-----LRFSVCLFRRIVACHFIKPRTLLDYW 296
             |   |     : :|:| :||::  | | |  :   |
Db     268 ----PFQNSGFNHWKMAYSLCRYRRVITVHQISPEEMHRIW 304


RESULT 6
Q91V52
ID   Q91V52      PRELIMINARY;      PRT;   305 AA.
AC   Q91V52;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase I (Fragment).
GN   B3GT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VARIOUS STRAINS;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039134; BAB68658.1; -.
DR   EMBL; AB039135; BAB68659.1; -.
DR   EMBL; AB039136; BAB68660.1; -.
DR   EMBL; AB039137; BAB68661.1; -.
DR   EMBL; AB039138; BAB68662.1; -.
DR   EMBL; AB039139; BAB68663.1; -.
DR   EMBL; AB039140; BAB68664.1; -.
DR   EMBL; AB039141; BAB68665.1; -.
DR   EMBL; AB039142; BAB68666.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1       1
FT   NON_TER     305     305
SQ   SEQUENCE   305 AA;  35611 MW;  A338019F3B06EF5A CRC64;

  Query Match             34.8%;  Score 576;  DB 11;  Length 305;
  Best Local Similarity   42.3%;  Pred. No. 3e-46;
  Matches  119;  Conservative   43;  Mismatches   97;  Indels   22;  Gaps    5;

Qy      23 FSMYSLNPFKEQSFVYKKDGNFLKLPDTDCRQTPPFLVLLVTSSHKQLAERMAIRQTWGK 82
           |      |    ||       ||      | :  ||||:|::::||:   | |||:|||
Db      39 FGNIRTRPINPHSF------EFLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGD 92

Qy      83 ERMVKGKQLKTFFLLGTTSSAAETKEVDQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWV 142
           |   || :: | ||||  :     : |:||||   ||| :||:| |:|||||:||: ||
Db      93 ENNFKGIKIATLFLLGKNADPVLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWV 152

Qy     143 HRFCPQAAFVMKTDSDMFINVDYLTELLLKKNR--TTRFFTGFLKLNEFPIRQPFSKWFV 200
             || :| :|||||||:|:|:| |   ||| :     |:|||:: :|  |||   |||::
Db     153 ATFCSKAKYVMKTDSDIFVNMDNLIYKLLKPSTKPRRRYFTGYV-INGGPIRDVRSKWYM 211

Qy     201 SKSEYPWDRYPPFCSGTGYVFSGDVASQVYNVSKSVPYIKLEDVFVGLCLERLNIRLEEL 260
           :  ||   ||||||||||:|| |||  :|  |     : ||||:||||| :| |
Db     212 PRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIH---- 267

Qy     261 HSQPTFFPGG-----LRFSVCLFRRIVACHFIKPRTLLDYW 296
             |   |     : :|:| :||::  | | |  :   |
Db     268 ----PFQNSGFNHWKMAYSLCRYRRVITVHQISPEEMHRIW 304


RESULT 7
O43825
ID   O43825      PRELIMINARY;      PRT;   422 AA.
AC   O43825;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE   2 (EC 2.4.1.-) (BETA3GALT2).
GN   GALT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   TISSUE=BRAIN;
RX   MEDLINE=98079080; PubMed=9417100;
RA   Kolbinger F., Streiff M.B., Katopodis A.G.;
RT   "Cloning of a human UDP-galactose:2-acetamido-2-deoxy-D-glucose 3beta-
RT   galactosyltransferase catalyzing the formation of type 1 chains.";
RL   J. Biol. Chem. 273:433-440(1998).
RN   [2]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE=98250717; PubMed=9582303;
RA   Amado M., Almeida R., Carneiro F., Levery S.B., Holmes E.H.,
RA   Nomoto M., Hollingsworth M.A., Hassan H., Schwientek T., Nielsen P.A.,
RA   Bennett E.P., Clausen H.;
RT   "A family of human beta3-galactosyltransferases. Characterization of
RT   four members of a UDP-galactose:beta-N-acetyl-glucosamine/beta-
RT   nacetyl-galactosamine beta-1,3-galactosyltransferase family.";
RL   J. Biol. Chem. 273:12770-12778(1998).
CC   -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC       COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC       WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT1 AND GALT2
CC       ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH DIFFERENT
CC       AFFINITY.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC       GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC       GLUCOSAMINYLGLYCOPEPTIDE.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN HEART AND TO A LESSER EXTENT IN
CC       BRAIN. NO EXPRESSION IN PLACENTA, KIDNEY, LIVER, SKELETAL MUSCLE,
CC       LUNG AND PANCREAS.
DR   EMBL; Y15014; CAA75245.1; -.
DR   EMBL; Y15060; CAA75344.1; -.
DR   MIM; 603018; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Glycoprotein; Golgi stack.
FT   DOMAIN        1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25     43       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       44    422       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    100    100       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   422 AA;  49213 MW;  BA62942D6AFB4B0C CRC64;

  Query Match             34.3%;  Score 569;  DB 4;  Length 422;
  Best Local Similarity   41.5%;  Pred. No. 2.1e-45;
  Matches  108;  Conservative   53;  Mismatches   95;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | |||||| | :  | |:   |||| :   :    : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
             :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
```

```
          ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db    266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy    228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
          ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db    326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy    288 KPRTLLDYWQALENSRGEDC 307
          :|  |: ||  |: ::   |
Db    386 QPSELIKYWNHLQQNKHNAC 405


RESULT 8
Q920V4
ID   Q920V4      PRELIMINARY;      PRT;   409 AA.
AC   Q920V4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CAST/EI;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039147; BAB68671.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47247 MW;  2425654F23A8F0E3 CRC64;

  Query Match             34.2%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 3.1e-45;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy     52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
          |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :      :
Db    140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy    110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
          :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db    200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy    170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
          ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db    260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy    228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
          ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db    320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy    288 KPRTLLDYWQALENSRGEDC 307
          :|  |: ||  |: ::   |
Db    380 QPSELIKYWNHLQQNKHNAC 399


RESULT 9
Q920V3
ID   Q920V3      PRELIMINARY;      PRT;   409 AA.
AC   Q920V3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HMI/MSF;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039151; BAB68675.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47220 MW;  2D2EBC4D68E8463D CRC64;

  Query Match             34.2%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 3.1e-45;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy     52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
          |::  |||:||: :   |:  | ||||||| | :  | |:   |||| :   :      :
Db    140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy    110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
          :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db    200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy    170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
          ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db    260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy    228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
          ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db    320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy    288 KPRTLLDYWQALENSRGEDC 307
          :|  |: ||  |: ::   |
Db    380 QPSELIKYWNHLQQNKHNAC 399


RESULT 10
Q920V2
ID   Q920V2      PRELIMINARY;      PRT;   409 AA.
AC   Q920V2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus spicilegus (Steppe mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZBN;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039153; BAB68677.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47383 MW;  0DC51AB6335A58DE CRC64;
```

```
  Query Match             34.2%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 3.1e-45;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | |||||||| | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
           :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db     320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     380 QPSELIKYWNHLQQNKHNAC 399


RESULT 11
Q91VE9
ID   Q91VE9      PRELIMINARY;      PRT;   409 AA.
AC   Q91VE9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BLG2/MSF, AND NJL/MSF;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039146; BAB68670.1; -.
DR   EMBL; AB039152; BAB68676.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47196 MW;  06D262B08B0B90F4 CRC64;

  Query Match             34.2%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 3.1e-45;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | |||||||| | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
           :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db     320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     380 QPSELIKYWNHLQQNKHNAC 399


RESULT 12
Q91V58
ID   Q91V58      PRELIMINARY;      PRT;   409 AA.
AC   Q91V58;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/10SNJ, BFM/2MSF, AND PGN2;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039144; BAB68668.1; -.
DR   EMBL; AB039145; BAB68669.1; -.
DR   EMBL; AB039149; BAB68673.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47233 MW;  9DBE67873273F13A CRC64;

  Query Match             34.2%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 3.1e-45;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | |||||||| | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
           :||::: ||||:::|| ||||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db     320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     380 QPSELIKYWNHLQQNKHNAC 399


RESULT 13
Q91V19
ID   Q91V19      PRELIMINARY;      PRT;   409 AA.
AC   Q91V19;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase II (Fragment).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSM/MSF, AND SWN/MSF;
RA   Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT   "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT   Five Mus musculus subspecies.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB039148; BAB68672.1; -.
DR   EMBL; AB039150; BAB68674.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     409    409
SQ   SEQUENCE   409 AA;  47206 MW;  D7BD514EE40A0157 CRC64;

  Query Match             34.2%;  Score 567;  DB 11;  Length 409;
  Best Local Similarity   41.2%;  Pred. No. 3.1e-45;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | |||||||  | :  | |:   |||| :   :      :
Db     140 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 199

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
            :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db     200 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 259

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     260 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 319

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db     320 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 379

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     380 QPSELIKYWNHLQQNKHNAC 399


RESULT 14
O54905
ID   O54905      PRELIMINARY;      PRT;   422 AA.
AC   O54905;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UDP-galactose:BETA-N-acetyl-glucosamine-BETA-1,3-galactosyltransferase
DE   II (EC 2.4.1.-) (UDP-GAL:BETAGLCNAC BETA 1,3-GALACTOSYLTRANFERASE-II)
DE   (BETA3GALT-II).
GN   B3GT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC   STRAIN=129/SVJ;
RX   MEDLINE=98079027; PubMed=9417047;
RA   Hennet T., Dinter A., Kuhnert P., Mattu T.S., Rudd P.M., Berger E.G.;
RT   "Genomic cloning and expression of three murine UDP-galactose: beta-N-
RT   acetylglucosamine beta1,3-galactosyltransferase genes.";
RL   J. Biol. Chem. 273:58-65(1998).
CC   -!- FUNCTION: THIS PROTEIN IS RESPONSIBLE FOR THE SYNTHESIS OF
CC       COMPLEX-TYPE N-LINKED OLIGOSACCHARIDES IN MANY GLYCOPROTEINS A AS
CC       WELL AS THE CARBOHYDRATE MOIETIES OF GLYCOLIPIDS. GALT-I, GALT-II
CC       AND GALT-III ENZYMES ACT ON THE SAME ACCEPTOR SUBSTRATES BUT WITH
CC       DIFFERENT AFFINITY.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETIL-BETA-D-GLUCOSAMINYL-
CC       GLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,3-N-ACETYL-BETA-D-
CC       GLUCOSAMINYLGLYCOPEPTIDE.
CC   -!- COFACTOR: MANGANESE.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND HEART. BARELY DETECTED
CC       IN OVARY, COLON AND LYMPH NODES. NO EXPRESSED IN TESTIS, UTERUS,
CC       STOMACH, KIDNEY, LIVER, LUNG, BONE MARROW, THYMUS AND SPLEEN.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-14 IS THE INITIATOR.
DR   EMBL; AF029791; AAC53524.1; -.
DR   MGD; MGI:1349461; B3galt2.
DR   InterPro; IPR002659; Galactosyl_T.
DR   Pfam; PF01762; Galactosyl_T; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW   Glycoprotein; Golgi stack; Manganese.
FT   DOMAIN        1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25     43       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       44    422       LUMENAL, CATALYTIC (POTENTIAL).
FT   CARBOHYD     75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   422 AA;  49107 MW;  D03D18A7484DCB03 CRC64;

  Query Match             34.2%;  Score 567;  DB 11;  Length 422;
  Best Local Similarity   41.2%;  Pred. No. 3.2e-45;
  Matches  107;  Conservative   53;  Mismatches   96;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | |||||||  | :  | |:   |||| :   :      :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRQTWGNETLAPGIQIIRVFLLGISIKLNGYLQHAI 205

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
            :||::: ||||:::|| |||||:||:||: ||  :||   :|||||||||:| :||
Db     206 QEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHTPYVMKTDSDMFVNTEYLIHK 265

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   |  :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| :      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRVDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::   |
Db     386 QPSELIKYWNHLQQNKHNAC 405


RESULT 15
Q9BZQ9
ID   Q9BZQ9      PRELIMINARY;      PRT;   422 AA.
AC   Q9BZQ9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   B3GALT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21218927; PubMed=11318611;
RA   Sood R., Bonner T.I., Malakowska I., Stephan D.A., Robbins C.M.,
RA   Connors T.D., Morgenbesser S.D., Su K., Faruque M.U., Pinkett H.,
RA   Graham C., Baxevanis A.D., Klinger K.W., Landes G.M., Trent J.M.,
RA   Carpten J.D.;
RT   "Cloning and characterization of 13 novel transcripts and the human
RT   RGS8 gene from the 1q25 region encompassing the hereditary prostate
RT   cancer (hpc1) locus.";
RL   Genomics 73:211-222(2001).
DR   EMBL; AF288390; AAG60610.1; -.
DR   InterPro; IPR002659; Galactosyl_T.
```

```
DR   Pfam; PF01762; Galactosyl_T; 1.
SQ   SEQUENCE   422 AA;  49241 MW;  BA6294E26AFB4ADC CRC64;

  Query Match             34.1%;  Score 565;  DB 4;  Length 422;
  Best Local Similarity   41.2%;  Pred. No. 4.9e-45;
  Matches  107;  Conservative   54;  Mismatches   95;  Indels    4;  Gaps    2;

Qy      52 CRQTPPFLVLLVTSSHKQLAERMAIRQTWGKERMVKGKQLKTFFLLGTT--SSAAETKEV 109
           |::  |||:||: :   |:  | |||:||| | :  | |:   |||| :   :    : :
Db     146 CQEKSPFLILLIAAEPGQIEARRAIRRTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAI 205

Qy     110 DQESQRHGDIIQKDFLDVYYNLTLKTMMGIEWVHRFCPQAAFVMKTDSDMFINVDYLTEL 169
            :||::: ||||:::|| ||||||:||:||: ||  :||   :||||||||:| :||
Db     206 LEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINK 265

Qy     170 LLKKNRTTR--FFTGFLKLNEFPIRQPFSKWFVSKSEYPWDRYPPFCSGTGYVFSGDVAS 227
           ||| :   | :|||:|     | |   |||::    || :||| ||||||||||||:|
Db     266 LLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAE 325

Qy     228 QVYNVSKSVPYIKLEDVFVGLCLERLNIRLEELHSQPTFFPGGLRFSVCLFRRIVACHFI 287
           ::: ||  :  : ||||:||:|| :| |      ::  |    : :| | :  ::  |
Db     326 KIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQF 385

Qy     288 KPRTLLDYWQALENSRGEDC 307
           :|  |: ||  |: ::  |
Db     386 QPSELIKYWNHLQQNKHNAC 405


Search completed: April 14, 2003, 12:09:05
Job time : 35 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2003, 14:50:38 ; Search time 154.708 Seconds
(without alignments)
13552.085 Million cell updates/sec

Title: US-09-914-152-2_COPY_401_1331
Perfect score: 931
Sequence: 1 aatggctttcccgaagatga.....gggaagattgtccgcctgtc 931

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	931	100.0	2775	21	AAA93875	Human beta-1,3 gal
2	930	99.9	10562	21	AAA93876	Human beta3Gal-T5
3	925.2	99.4	1011	21	AAA27959	Human beta3Gal-T5
4	163.4	17.6	1739	15	AAQ67067	Beta-1,3-galactosy
5	130.4	14.0	1773	20	AAX87193	Human Dendriac cDN
6	130.4	14.0	1897	24	ABK51201	Human cDNA encodin
7	130.4	14.0	2095	21	AAZ65022	Membrane-bound pro
8	130.4	14.0	2095	22	AAS46005	Human DNA encoding
9	130.4	14.0	2095	22	AAF92075	Human PRO1074 cDNA
10	130.4	14.0	2095	22	AAF44168	Human PRO1074 (UNQ
11	130.4	14.0	2168	22	AAH15711	Human cDNA sequenc
12	130.4	14.0	2189	20	AAX87192	Human Dendriac cDN
13	122.4	13.1	1266	20	AAX35710	cDNA encoding a pr
14	122.4	13.1	2420	20	AAX35711	cDNA encoding a pr
15	114.8	12.3	1134	21	AAA58791	DNA encoding the b
16	108.2	11.6	1134	22	AAD09356	Human beta-1,3-gal
17	93.2	10.0	1725	23	ABL14319	Drosophila melanog
18	93.2	10.0	4218	23	ABL14318	Drosophila melanog
19	89.8	9.6	1116	19	AAV49599	Human epidermoid c
20	89.8	9.6	1208	24	AAS16945	Human beta1,3-N-ac
21	89.8	9.6	2180	22	AAF29257	Human beta 1,3-N-a
22	89.8	9.6	2186	19	AAV49598	Human epidermoid c
23	89.8	9.6	2186	24	ABL41994	Nucleotide sequenc
24	89.8	9.6	2198	22	AAF93831	Human cDNA encodin
25	89.8	9.6	2205	22	AAF29256	Human beta 1,3-N-a
26	89.8	9.6	2210	22	AAS46068	Human DNA encoding
27	89.8	9.6	2427	23	ABV22526	Human prostate exp
28	89.8	9.6	2427	23	ABV28341	Human prostate exp
29	87.6	9.4	1337	24	AAS16946	Murine beta1,3-N-a
30	83.4	9.0	3033	23	ABL14314	Drosophila melanog
31	83.4	9.0	3330	23	ABL14322	Drosophila melanog
32	81.8	8.8	1271	20	AAX97916	Human secreted pro
33	81.8	8.8	1271	20	AAX87194	Human Brainiac-3 c
34	78.4	8.4	1296	22	AAF29258	Human beta 1,3-N-a
35	78.4	8.4	1420	21	AAA58790	DNA encoding the b
36	78.4	8.4	1642	21	AAC58115	Human PRO4344 nucl
37	78.4	8.4	1643	24	ABK69978	cDNA encoding huma
38	77.8	8.4	1359	23	AAS84438	DNA encoding novel
39	77.8	8.4	1360	24	AAD35225	Human TRNFR-10 cDN
40	74.6	8.0	977	21	AAA57383	DNA encoding a Bra
41	74.6	8.0	2117	24	AAD35219	Human TRNFR-4 cDNA
42	74.6	8.0	2684	21	AAC58117	Human PRO4397 nucl
43	74.6	8.0	2684	24	ABK69977	cDNA encoding huma
44	70.2	7.5	400	21	AAZ94080	Haematopoietic ste
45	68.6	7.4	1191	21	AAA53023	Human beta-1,3-gal

ALIGNMENTS

RESULT 1
AAA93875
ID AAA93875 standard; DNA; 2775 BP.
XX
AC AAA93875;
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta-1,3 galactose transferase encoding DNA.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; ds.
XX
OS Homo sapiens.
XX
PN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI; 2000-549409/50.
DR P-PSDB; AAB93875.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of

PT digestive system cancer -
XX
PS Claim 5; Page 99-102; 123pp; Japanese.
XX
CC This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAcbeta1-3Galbeta1-4Glc) to give Galbeta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents Beta-1,3 galactose
CC transferase encoding DNA.
XX
SQ Sequence 2775 BP; 681 A; 698 C; 669 G; 727 T; 0 other;

Query Match 100.0%; Score 931; DB 21; Length 2775;
Best Local Similarity 100.0%; Pred. No. 6.7e-278;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 AATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTG 460

Qy      61 TTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 TTTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAA 520

Qy     121 AGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 AGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGT 580

Qy     181 CCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 CCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTG 640

Qy     241 GGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 GGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCAC 700

Qy     301 CAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTAT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     701 CAGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTAT 760

Qy     361 CCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     761 CCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGA 820

Qy     421 ATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     821 ATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTT 880

Qy     481 CATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     881 CATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTT 940

Qy     541 CACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     941 CACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGT 1000

Qy     601 CAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1001 CAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGT 1060

Qy     661 GTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1061 GTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAA 1120

Qy     721 ACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1121 ACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCT 1180

Qy     781 CCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1181 CCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAG 1240

Qy     841 GATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1241 GATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGA 1300

Qy     901 GAATTCCCGGGGGGAAGATTGTCCGCCTGTC 931
          |||||||||||||||||||||||||||||||
Db    1301 GAATTCCCGGGGGGAAGATTGTCCGCCTGTC 1331
```

RESULT 2
AAA93876
ID AAA93876 standard; DNA; 10562 BP.
XX
AC AAA93876;
XX
DT 15-JAN-2001 (first entry)
XX
DE Human beta3Gal-T5 encoding DNA.
XX
KW Beta-1,3 galactose transferase; treatment; diagnosis; cancer; human;
KW digestive system; beta3Gal-T5; ds.
XX
OS Homo sapiens.
XX
PN WO200050608-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-JP01070.
XX
PR 25-FEB-1999; 99JP-0047571.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Narimatsu H, Isshiki S, Togayachi A, Sasaki K;
XX
DR WPI; 2000-549409/50.
XX
PT Beta-1,3 galactose transferase and DNA encoding it, useful for
PT synthesis of type 1 sialyl Lewis, a carbohydrate for treatment of
PT digestive system cancer -
XX
PS Claim 31; Page 103-111; 123pp; Japanese.
XX
CC This invention relates to a polypeptide (I) with beta-1,3 galactose
CC transferase activity, or variants of (I) comprising amino acid additions,
CC deletions and/or substitutions. Included in the invention is DNA encoding
CC all or part of (I); expression vectors containing the DNA, host cells
CC transformed by the vectors; a method for the preparation of the
CC polypeptide by culture of the transformants or by expression in the milk
CC of a transgenic mammal, and antibodies recognising (I). The Beta-1,3
CC galactose transferase protein transfers galactose by beta-1,3 bonding to
CC N-acetylglucosamine present in a non-cyclic carbohydrate chain (such as
CC GlcNAcbeta1-3Galbeta1-4Glc) to give Galbeta1-3GlcNAc. The protein and
CC DNA encoding it are useful for the treatment and diagnosis of cancer of
CC the digestive system. The present sequence represents a Beta3Gal-T5
CC encoding DNA sequence.
XX
SQ Sequence 10562 BP; 2610 A; 2415 C; 2574 G; 2963 T; 0 other;

Query Match 99.9%; Score 930; DB 21; Length 10562;
Best Local Similarity 100.0%; Pred. No. 2.9e-277;
Matches 930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2 ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGT 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    8234 ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGT 8293

Qy      62 TTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8294 TTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 8353

Qy     122 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8354 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 8413

Qy     182 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8414 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 8473

Qy     242 GGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8474 GGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 8533

Qy     302 AGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8534 AGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 8593

Qy     362 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8594 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 8653

Qy     422 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTC 481
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8654 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTC 8713

Qy     482 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 541
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8714 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 8773

Qy     542 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 601
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8774 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 8833

Qy     602 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8834 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 8893

Qy     662 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8894 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 8953

Qy     722 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8954 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 9013

Qy     782 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9014 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 9073

Qy     842 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9074 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 9133

Qy     902 AATTCCCGGGGGGAAGATTGTCCGCCTGTC 931
           |||||||||||||||||||||||||||||
Db    9134 AATTCCCGGGGGGAAGATTGTCCGCCTGTC 9163
```

```
RESULT 3
AAA27959
ID   AAA27959 standard; DNA; 1011 BP.
XX
AC   AAA27959;
XX
DT   15-AUG-2000  (first entry)
XX
DE   Human beta3Gal-T5 gene sequence.
XX
KW   UDP-D-galactose; beta-N-acetylglucosamine beta1,3-galactosyltransferase;
KW   Beta3Gal-T5; ss; human; chromosome 21q22.3; galactosylation;
KW   beta1,3-galactosyl glycosylated saccharide production; glycopeptide;
KW   glycoprotein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..1011
FT                   /*tag= a
FT                   /product= "Beta3Gal-T5"
FT   primer_bind     complement (79..98)
FT                   /*tag= b
FT   primer_bind     complement (150..170)
FT                   /*tag= c
FT   primer_bind     991..1011
FT                   /*tag= d
XX
PN   WO200029558-A1.
XX
PD   25-MAY-2000.
XX
PF   11-NOV-1999;   99WO-US26807.
XX
PR   13-NOV-1998;   98DK-0001483.
XX
PA   (CLAU/) CLAUSEN H.
XX
PI   Clausen H,  Amado M;
XX
DR   WPI; 2000-399728/34.
DR   P-PSDB; AAY94641.
XX
PT   Novel nucleic acid sequence encoding human
PT   UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
PT   useful for obtaining beta 1,3-galactosyl glycosylated saccharides and
PT   glycopeptides or glycoproteins  -
XX
PS   Claim 7; Fig 1; 74pp; English.
XX
CC   The present invention relates to a nucleic acid sequence encoding
CC   UDP-D-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase
CC   (beta3Gal-T5). Beta3 transferases add galactose to the hydroxy group at
CC   carbon 3 of 2-acetamido-2-deoxy-D-glucose (GlcNAc). The present sequence
CC   represents the human beta3Gal-T5 gene sequence. The beta3Gal-T5 gene is
CC   located on human chromosome 21q22.3. Beta3Gal-T5 is a type II
CC   transmembrane glycoprotein. The invention also relates to the beta3Gal-T5
CC   protein sequence, a nucleic acid vector comprising the beta3Gal-T5
CC   nucleotide sequence, a host cell comprising the vector, and a method for
CC   the production of the beta3Gal-T5 protein from the host cells. The
CC   methods of the invention can be used for recombinant production of
CC   beta3Gal-T5 for use as a catalyst and for recombinant production of
CC   peptides or proteins with appropriate galactosylation. The beta3Gal-T5
CC   protein can be used to obtain beta1,3-galactosyl glycosylated
CC   saccharides, glycopeptides or glycoproteins.
XX
SQ   Sequence 1011 BP; 247 A; 256 C; 251 G; 257 T; 0 other;
```

```
  Query Match             99.4%;  Score 925.2;  DB 21;  Length 1011;
  Best Local Similarity   99.7%;  Pred. No. 2.4e-276;
  Matches  927;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ATGGCTTTCCCGAAGATGAGATTGATGTATATTTGCCTTCTGGTTCTGGGGGCTCTTTGT 61
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      79 ATGGCTTTCCCGAAGATGAGATTGATGTATATCTGCCTTCTGGTTCTGGGGGCTCTTTGT 138

Qy      62 TTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 TTGTATTTTAGCATGTACAGTCTAAATCCTTTCAAAGAACAGTCCTTTGTTTACAAGAAA 198

Qy     122 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 GACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTC 258
```

```
Qy    182 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 CTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGG 318

Qy    242 GGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 301
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 GGGAAAGAGAGGACGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACC 378

Qy    302 AGCAGTGCAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 361
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    379 AGCAGTGCAGCGGAAACAAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATC 438

Qy    362 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 CAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAA 498

Qy    422 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTC 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 TGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTC 558

Qy    482 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    559 ATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAAACAGAACAACCAGGTTTTTC 618

Qy    542 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 ACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTC 678

Qy    602 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    679 AGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTG 738

Qy    662 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 721
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    739 TTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAA 798

Qy    722 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 781
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    799 CTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTC 858

Qy    782 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    859 CACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGG 918

Qy    842 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 901
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    919 ATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGAG 978

Qy    902 AATTCCCGGGGGGAAGATTGTCCGCCTGTC 931
          |||||||||||||||||||||||||||||
Db    979 AATTCCCGGGGGGAAGATTGTCCGCCTGTC 1008
```

RESULT 4
AAQ67067
ID AAQ67067 standard; cDNA to mRNA; 1739 BP.
XX
AC AAQ67067;
XX
DT 14-MAR-1995 (first entry)
XX
DE Beta-1,3-galactosyltransferase cDNA.
XX
KW Beta-1,3-galactosyltransferase; enzyme; saccharide chain; pAMoPRWM1;
KW KJM-1 cells; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 676..1656
FT /*tag= a
FT /product= Beta-1,3-galactosyltransferase
XX
PN JP06181759-A.
XX
PD 05-JUL-1994.
XX
PF 16-DEC-1992; 92JP-0336436.
XX
PR 16-DEC-1992; 92JP-0336436.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR WPI; 1994-251683/31.
DR P-PSDB; AAR57433.
XX
PT Beta-galactosyl-transferase DNA and protein - useful for prodn. of
PT saccharide chains
XX
PS Claim 3; Page 22-24; 47pp; Japanese.
XX
CC This sequence encodes a beta-1,3-galactosyltransferase. This enzyme
CC can be used to produce physiologically active saccharide chains and
CC varients, and for improvement of saccharide chains bound to
CC physiologically active proteins. This cDNA represents a fragment
CC of the plasmid pAMoPRWM1 which was cloned in KJM-1 cells.
XX
SQ Sequence 1739 BP; 473 A; 407 C; 422 G; 437 T; 0 other;

Query Match 17.6%; Score 163.4; DB 15; Length 1739;
Best Local Similarity 55.7%; Pred. No. 8.4e-40;
Matches 358; Conservative 0; Mismatches 276; Indels 9; Gaps 2;

```
Qy    129 ACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGG 188
          | || ||||   |      |   | ||   | | |    ||| || || ||  | ||
Db    866 AATTTCTTATCAACGAGCCCAATAAATGTGAGAAAAACATTCCTTTTCTTGTTATCCTCA 925

Qy    189 TGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAG 248
          | | | |  | |||||  | || | ||  ||   ||| ||| |  ||||||||||| | |
Db    926 TCAGCACCACTCACAAGGAATTTGATGCCCGTCAGGCAATCAGAGAGACGTGGGGGGATG 985

Qy    249 AGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCCTGGGGACCACCAGCAGTG 308
          |||  |   | |||||  |   || |   ||  | ||||||||||| |  |       |
Db    986 AGAACAACTTTAAGGGGATCAAGATAGCCACCCTGTTCCTCCTGGGCAAGAATGCTGATC 1045

Qy    309 CAGCGGAAACGAAAGAGGTGGACCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGG 368
          | |     |   |  |||||| || ||||||||      |   || || |||  | |||
Db    1046 CTGTTCTCAATCAGATGGTGGAGCAAGAGAGCCAAATCTTCCATGATATCATCGTGGAGG 1105

Qy    369 ATTTCCTAGACGTCTATTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCC 428
          | ||  | |||  |||  | || || |||||  || ||  | |||||  ||   ||||||
Db    1106 ACTTTATTGACTCCTACCATAACCTTACCCTCAAAACATTAATGGGGATGAGATGGGTGG 1165

Qy    429 ATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATG 488
              ||||||  |  | ||   || ||| ||||||||||||   |||||  ||  | ||
Db    1166 CCACTTTTTGTTCAAAAGCCAAGTATGTCATGAAAACAGACAGCGACATTTTTGTAAACA 1225

Qy    489 TTGACTATCTGACTGAACTGCTTCTGAA------GAAAAACAGAACAACCAGGTTTTTCA 542
          | ||| |||| | | |     | |||||       |  ||   |  ||  |||| ||| |
Db    1226 TGGACAATCTTATTTATAAATTACTGAAACCCTCCACCAAGCCACGAAGAAGGTATTTTA 1285

Qy    543 CTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCA 602
          ||||||   | |   | ||||     || ||  || |      ||| |||||||ｌ | |
Db    1286 CTGGCTATGTCA---TTAATGGAGGACCGATTCGGGATGTCCGCAGTAAGTGGTATATGC 1342

Qy    603 GTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGT 662
            |   |  || ||      |  |||||||| ||||| || || || |||||| | |
Db    1343 CCAGGGATTTGTACCCAGACAGTAACTACCCACCTTTCTGTTCGGGGACTGGCTACATCT 1402

Qy    663 TTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAAC 722
          |||| | ||| || ||    |  | |||||  |||      |          || | |
Db    1403 TTTCAGCCGATGTAGCTGAACTCATTTACAAGACCTCACTCCACACAAGGCTGCTTCACC 1462

Qy    723 TGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGAACAT 765
          | |||||||| | |||||| || || ||   || ||||  |||
```

```
Db   1463 TTGAAGACGTATATGTGGGACTGTGTCTTCGAAAGCTGGGCAT 1505


RESULT 5
AAX87193
ID   AAX87193 standard; cDNA; 1773 BP.
XX
AC   AAX87193;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Human Dendriac cDNA.
XX
KW   Dendriac; Brainiac-2; human; Notch; immune disorder;
KW   neurological disorder; diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..980
FT                   /*tag=  a
FT   sig_peptide     20..94
FT                   /*tag=  b
FT   mat_peptide     95..977
FT                   /*tag=  c
XX
PN   WO9931116-A1.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-US27049.
XX
PR   17-NOV-1998;   98US-0108928.
PR   18-DEC-1997;   97US-0068006.
PR   12-MAR-1998;   98US-0077687.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ebner R,  Endress GA,  Florence KA,  Rosen CA,  Ruben SM;
PI   Soppet DR,  Yu G;
XX
DR   WPI; 1999-430031/36.
DR   P-PSDB; AAY06461.
XX
PT   Human proteins for treating and detecting immune and neurological
PT   disorders
XX
PS   Claim 2; Page 119-120; 132pp; English.
XX
CC   This is the nucleotide sequence of a cDNA clone which codes for
CC   human Dendriac (see AAY06461), also called Brainiac-2, a novel
CC   member of the Brainiac family.  Dendriac cDNA (see also AAX87193) is
CC   deposited as ATCC 203056 and ATCC 209627.  It was initially
CC   discovered in a dendritic cell cDNA library.  Additional clones of
CC   the same gene were identified in NTERA2 cell, adult pulmonary
CC   tissue, salivary gland, ovary, Caco-2 colon adenocarcinoma, smooth
CC   muscle, cerebellum, 8-week-old human mrbyo, haemagiopericytoma,
CC   amygdala, substantia nigra and whole brain cDNA libraries.  Dendriac
CC   nucleic acids are used in the recombinant production of Dendriac
CC   polypeptides, especially the mature protein or epitope-bearing
CC   fragments.  Dendriac and Brainiac proteins are members of the Notch
CC   family of proteins involved in the control of cell division.  The
CC   proteins may be used to detect and/or treat immune system and
CC   neurological disorders.
XX
SQ   Sequence 1773 BP; 568 A; 300 C; 339 G; 566 T; 0 other;

  Query Match             14.0%;  Score 130.4;  DB 20;  Length 1773;
  Best Local Similarity   48.9%;  Pred. No. 1.5e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170
          ||||||   |||||    |||||      |      || || |||||   ||   |  ||

Db    157 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 216

Qy    171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230
          | || || ||| | ||||||||||||    ||     | |  ||   |   |   |||||||
Db    217 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 276

Qy    231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290
          |    || |||||  || | | |      |  |||| |  || |    |||||  ||| |
Db    277 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 336

Qy    291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341
          | ||             |  || |  |   ||             | ||  | ||   ||
Db    337 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 396

Qy    342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401
          |   | || |||||  ||||   | |||||  |||||   ||| | ||  |||||| |||
Db    397 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 456

Qy    402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461
          | |||||  ||||   |    |||||   |   |||||  ||  | ||   ||  || ||||
Db    457 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 516

Qy    462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521
          | |||||| | ||  | |||||||||  || | || |      |    ||| | ||   ||
Db    517 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 576

Qy    522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581
          ||    ||   | |||||||||| || |        | |  || | | | ||   ||
Db    577 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 636

Qy    582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641
          | |   ||      | |  | |    || ||||| |   |   || ||| |||| ||
Db    637 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 696

Qy    642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    697 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 756

Qy    702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761
            |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    757 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 816

Qy    762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db    817 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 876

Qy    822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881
          || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    877 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 936

Qy    882 ACTACTGGCAGG 893
            |  |||||||
Db    937 CTTTTTGGCAGG 948


RESULT 6
ABK51201
ID   ABK51201 standard; cDNA; 1897 BP.
XX
AC   ABK51201;
XX
DT   30-JUL-2002  (first entry)
XX
DE   Human cDNA encoding beta1,3-acetylgalactosamine transferase.
XX
KW   Human; ss; gene; beta1,3-acetylgalactosamine transferase;
KW   Gb4 sugar chain.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    109..1101
FT                   /*tag=  a
```

```
FT                   /note= "This region is specifically claimed in claim 1"
FT   CDS             109..1104
FT                   /*tag= b
FT                   /product= "Beta1,3-acetylgalactosamine transferase"
FT   polyA_signal    1157..1162
FT                   /*tag= c
XX
PN   JP2002085069-A.
XX
PD   26-MAR-2002.
XX
PF   08-SEP-2000; 2000JP-0273835.
XX
PR   08-SEP-2000; 2000JP-0273835.
XX
PA   (SEGK ) SEIKAGAKU KOGYO CO LTD.
PA   (FURU/) FURUKAWA K.
XX
DR   WPI; 2002-378274/41.
DR   P-PSDB; AAU80224.
XX
PT   Preparation of beta-1,3-acetylgalactosamine transferase -
XX
PS   Claim 1; Page 11-13; 15pp; Japanese.
XX
CC   The invention relates to the preparation of beta-1,3-acetylgalactosamine
CC   (a Gb4 sugar chain molecule) comprising introducing a DNA appearing as
CC   ABK51201 (S1) or a DNA hybridising with S1 or a base sequence
CC   complementary to the base sequence or part of these base sequences under
CC   a stringent condition into a cell and growing the cell to express beta-1,
CC   3-acetylgalactosamine transferase and collecting it. Also included is the
CC   preparation of Gb4 sugar chain comprising contacting a polypeptide
CC   comprising the beta-1,3-acetylgalactosamine transferase protein appearing
CC   as AAU80224 (A) or a polypeptide consisting of an amino acid sequence in
CC   which at least one amino acid is replaced, deleted, inserted or
CC   transferred in the amino acid sequence (A) and having enzymatic activity
CC   transferring N-acetylgalactosamine residue from an N-acetylgalactosamine
CC   donor to the C3 site of the galactose residue in Gb3 sugar chain which is
CC   the receptor with an N-acetylgalactosamine donor and Gb3 sugar chain.
CC   The present sequence is the cDNA encoding beta-1,3-acetylgalactosamine
CC   transferase.
XX
SQ   Sequence 1897 BP; 596 A; 327 C; 366 G; 608 T; 0 other;

  Query Match             14.0%;  Score 130.4;  DB 24;  Length 1897;
  Best Local Similarity   48.9%;  Pred. No. 1.5e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170
          ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db    281 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 340

Qy    171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230
          | || || ||| | ||||||||||   ||    | |  ||   |   |   ||||||
Db    341 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 400

Qy    231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290
          |    || |||||  || | | |    |  |||| |  || |    |||||  ||| |
Db    401 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 460

Qy    291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341
          | ||                |  || |  |    ||         | ||  | ||   ||
Db    461 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 520

Qy    342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401
            |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db    521 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 580

Qy    402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461
          | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db    581 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 640

Qy    462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521
          | |||||| | ||  | |||||||||  || | || |     |    ||| | ||   ||
Db    641 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 700

Qy    522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    701 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 760

Qy    582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641
            | |   ||       | |   | |    || ||||| |   |   || ||| |||| ||
Db    761 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 820

Qy    642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    821 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 880

Qy    702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761
            |||   |  ||| ||  | ||||| || | ||| ||| |||| |  |    | |
Db    881 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 940

Qy    762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821
          |  | |   |  |    |   |   |          || |||| |    |  | | ||
Db    941 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1000

Qy    822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881
          || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db   1001 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1060

Qy    882 ACTACTGGCAGG 893
            |  |||||||
Db   1061 CTTTTTGGCAGG 1072


RESULT 7
AAZ65022
ID   AAZ65022 standard; cDNA; 2095 BP.
XX
AC   AAZ65022;
XX
DT   05-APR-2000  (first entry)
XX
DE   Membrane-bound protein PRO1074 encoding cDNA.
XX
KW   Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW   pharmaceutical; receptor immunoadhesin; gene mapping; ss.
XX
OS   Homo sapiens.
XX
PN   WO9963088-A2.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999;   99WO-US12252.
XX
PR   02-JUN-1998;   98US-0087607.
PR   02-JUN-1998;   98US-0087609.
PR   02-JUN-1998;   98US-0087759.
PR   03-JUN-1998;   98US-0087827.
PR   04-JUN-1998;   98US-0088021.
PR   04-JUN-1998;   98US-0088025.
PR   04-JUN-1998;   98US-0088028.
PR   04-JUN-1998;   98US-0088029.
PR   04-JUN-1998;   98US-0088030.
PR   04-JUN-1998;   98US-0088033.
PR   04-JUN-1998;   98US-0088326.
PR   05-JUN-1998;   98US-0088167.
PR   05-JUN-1998;   98US-0088202.
PR   05-JUN-1998;   98US-0088212.
PR   05-JUN-1998;   98US-0088217.
PR   09-JUN-1998;   98US-0088655.
PR   10-JUN-1998;   98US-0088722.
PR   10-JUN-1998;   98US-0088730.
PR   10-JUN-1998;   98US-0088734.
PR   10-JUN-1998;   98US-0088738.
```

PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
DR WPI; 2000-072883/06.
DR P-PSDB; AAY66685.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
PS Claim 2; Fig 136; 822pp; English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
SQ Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

```
  Query Match             14.0%;  Score 130.4;  DB 21;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 1.6e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170
           ||||||   |||||    |||||       |      || || |||||   ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230
           | || || ||| | ||||||||||   ||    | |  ||   |   |   ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290
           |    || |||||  || | | |     |  |||| |  || |    ||||| ||| |
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341
           | ||                |  || | |    ||         | ||  | ||   ||
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401
             |   | || ||||| ||||   | |||||  |||||   ||| | || ||||||  |||
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461
           | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521
           | |||||| | ||  | ||||||||||  || | || |     |    ||| | ||   ||
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581
           ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641
             | |  ||       | |   | |    || ||||| |    |   || ||| |||| ||
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy     642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701
           ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy     702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy     762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821
           |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy     822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881
             || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy     882 ACTACTGGCAGG 893
             |  |||||||
Db    1274 CTTTTTGGCAGG 1285


RESULT 8
AAS46005
ID   AAS46005 standard; cDNA; 2095 BP.
XX
AC   AAS46005;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human DNA encoding PRO polypeptide sequence #81.
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW   dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168848-A2.
XX
PD   20-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06520.
XX
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   03-MAR-2000; 2000US-187202P.
PR   06-MAR-2000; 2000US-186968P.
PR   14-MAR-2000; 2000US-189320P.
PR   14-MAR-2000; 2000US-189328P.
PR   15-MAR-2000; 2000WO-US06884.
PR   21-MAR-2000; 2000US-190828P.
PR   21-MAR-2000; 2000US-191007P.
PR   21-MAR-2000; 2000US-191048P.
PR   21-MAR-2000; 2000US-191314P.
PR   28-MAR-2000; 2000US-192655P.
PR   29-MAR-2000; 2000US-193032P.
PR   29-MAR-2000; 2000US-193053P.
PR   30-MAR-2000; 2000WO-US08439.
PR   04-APR-2000; 2000US-194449P.
PR   04-APR-2000; 2000US-194647P.
PR   11-APR-2000; 2000US-195975P.
PR   11-APR-2000; 2000US-196000P.
PR   11-APR-2000; 2000US-196187P.
PR   11-APR-2000; 2000US-196690P.
PR   11-APR-2000; 2000US-196820P.
PR   18-APR-2000; 2000US-198121P.
PR   18-APR-2000; 2000US-198585P.
PR   25-APR-2000; 2000US-199397P.
PR   25-APR-2000; 2000US-199550P.
PR   25-APR-2000; 2000US-199654P.
PR   03-MAY-2000; 2000US-201516P.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   05-JUN-2000; 2000US-209832P.
PR   28-JUL-2000; 2000WO-US20710.
PR   22-AUG-2000; 2000US-0644848.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI   Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2001-602746/68.
DR   P-PSDB; AAU29104.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
XX
PS   Claim 2; Fig 161; 774pp; English.
XX
CC   Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC   primers for PRO polypeptides of the invention. The sequences of the
CC   invention can be used to detect the presence of a tumour in a mammal by
CC   comparing the level of expression of a PRO polypeptide in a test sample
CC   of cells from the animal and a control sample of normal cells, whereby a
CC   higher level of expression in the test sample indicates the presence of a
CC   tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
```

```
CC   pigs, goats and rabbits but are preferably human. The polypeptides can be
CC   used to stimulate tumour necrosis factor (TNF) alpha release from human
CC   blood, when contacted with it. A specific polypeptide can be used to
CC   stimulate the proliferation or differentiation of chondrocyte cells. The
CC   PRO proteins can be used to determine the presence of tumours and also
CC   susceptibility to tumour development, particularly adrenal, lung, colon,
CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC   can be used for genetic analysis of individuals with genetic disorders.
XX
SQ   Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

  Query Match             14.0%;  Score 130.4;  DB 22;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 1.6e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170
           ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230
           | || || ||| | ||||||||||    ||    | |  ||   |   |   ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290
           |    || |||||  || | | |      |  |||| |  || |    ||||| ||| |
Db     614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy     291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341
           | ||                 |  || |  |    ||          | ||  | ||   ||
Db     674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy     342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401
           |   | || ||||| ||||   | |||||  |||||   ||| | || |||||| |||
Db     734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy     402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461
           | ||||| ||||   |    |||||   |   ||||| ||  | ||   ||  || ||||
Db     794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy     462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521
           | |||||| | ||  | ||||||||||  || | || |      |    ||| | ||   ||
Db     854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy     522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581
           ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db     914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy     582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641
           | |   ||     | |   | |    || ||||| |   |   || ||| |||| ||
Db     974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy     642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701
           ||   ||    || ||  |  | ||  | ||  ||| |      | | ||  |  |
Db    1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy     702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761
             |||   |  ||| ||  | ||||| || | ||| ||| ||||  |  |     | |
Db    1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy     762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821
           |  | |   |  |    |   |    |        || |||| |    |  | |  ||
Db    1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy     822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881
             || || |   | ||  |  |  | ||   ||| | |      |   |    || |
Db    1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy     882 ACTACTGGCAGG 893
           |  |||||||
Db    1274 CTTTTTGGCAGG 1285
```

```
RESULT 9
AAF92075
ID   AAF92075 standard; cDNA; 2095 BP.
XX
AC   AAF92075;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human PRO1074 cDNA.
XX
KW   Human; PRO protein; mapping; ss.
XX
OS   Homo sapiens.
XX
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000;  2000US-0175481.
PR   18-FEB-2000;  2000WO-US04341.
PR   18-FEB-2000;  2000WO-US04342.
PR   22-FEB-2000;  2000WO-US04414.
PR   01-MAR-2000;  2000WO-US05601.
PR   03-MAR-2000;  2000US-0187202.
PR   25-APR-2000;  2000US-0199397.
PR   22-MAY-2000;  2000WO-US14042.
PR   05-JUN-2000;  2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   P-PSDB; AAB87543.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 2; Fig 35; 278pp; English.
XX
CC   The present sequence is the coding sequence for a human PRO polypeptide
CC   (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
CC   antagonists or anti-PRO antibodies are useful for preparation of a
CC   medicament useful in the treatment of a condition which is responsive to
CC   the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
CC   protein may also be employed as molecular weight markers for protein
CC   electrophoresis. The PRO coding sequence has applications in molecular
CC   biology, including use as hybridisation probes, and in chromosome and
CC   gene mapping.
XX
SQ   Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

  Query Match             14.0%;  Score 130.4;  DB 22;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 1.6e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy     111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170
           ||||||   |||||    |||||      |      || || |||||   ||   |  ||
Db     494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy     171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230
           | || || ||| | ||||||||||    ||    | |  ||   |   |   ||||||
Db     554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy     231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290
```

```
Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341

Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401

Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461

Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521

Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581

Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641

Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701

Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761

Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821

Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881

Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    882 ACTACTGGCAGG 893

Db   1274 CTTTTTGGCAGG 1285


RESULT 10
AAF44168
ID   AAF44168 standard; cDNA; 2095 BP.
XX
AC   AAF44168;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO1074 (UNQ531) nucleotide sequence SEQ ID NO:208.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   P-PSDB; AAB65208.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 2; Fig 136; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence 2095 BP; 629 A; 382 C; 418 G; 666 T; 0 other;

  Query Match          14.0%;  Score 130.4;  DB 22;  Length 2095;
  Best Local Similarity   48.9%;  Pred. No. 1.6e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170

Db    494 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 553

Qy    171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230

Db    554 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 613

Qy    231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290

Db    614 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 673

Qy    291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341
```

```
Db    674 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 733

Qy    342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401

Db    734 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 793

Qy    402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461

Db    794 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 853

Qy    462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521

Db    854 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 913

Qy    522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581

Db    914 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 973

Qy    582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641

Db    974 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1033

Qy    642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701

Db   1034 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1093

Qy    702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761

Db   1094 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1153

Qy    762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821

Db   1154 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1213

Qy    822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881

Db   1214 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1273

Qy    882 ACTACTGGCAGG 893

Db   1274 CTTTTTGGCAGG 1285


RESULT 11
AAH15711
ID   AAH15711 standard; cDNA; 2168 BP.
XX
AC   AAH15711;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:14092.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14092; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2168 BP; 634 A; 405 C; 452 G; 677 T; 0 other;

  Query Match           14.0%;  Score 130.4;  DB 22;  Length 2168;
  Best Local Similarity  48.9%;  Pred. No. 1.7e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170

Db    571 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 630

Qy    171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230

Db    631 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 690

Qy    231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290

Db    691 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 750

Qy    291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341

Db    751 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 810

Qy    342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401

Db    811 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 870

Qy    402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461

Db    871 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAACGCCAAGTACGTAATGA 930

Qy    462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521

Db    931 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 990

Qy    522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581

Db    991 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 1050

Qy    582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641
```

```
Db   1051 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTCCCCTCCATACT 1110

Qy    642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1111 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1170

Qy    702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761
            |||   |  ||| ||  | ||||| || | ||| ||| |||| |  |     | |
Db   1171 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1230

Qy    762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db   1231 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1290

Qy    822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881
          || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db   1291 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1350

Qy    882 ACTACTGGCAGG 893
            |  |||||||
Db   1351 CTTTTTGGCAGG 1362


RESULT 12
AAX87192
ID   AAX87192 standard; cDNA; 2189 BP.
XX
AC   AAX87192;
XX
DT   27-SEP-1999  (first entry)
XX
DE   Human Dendriac cDNA.
XX
KW   Dendriac; Brainiac-2; human; Notch; immune disorder;
KW   neurological disorder; diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             435..1394
FT                   /*tag= a
FT   sig_peptide     435..509
FT                   /*tag= b
FT   mat_peptide     510..1391
FT                   /*tag= c
XX
PN   WO9931116-A1.
XX
PD   24-JUN-1999.
XX
PF   17-DEC-1998;   98WO-US27049.
XX
PR   17-NOV-1998;   98US-0108928.
PR   18-DEC-1997;   97US-0068006.
PR   12-MAR-1998;   98US-0077687.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ebner R,  Endress GA,  Florence KA,  Rosen CA,  Ruben SM;
PI   Soppet DR,  Yu G;
XX
DR   WPI; 1999-430031/36.
DR   P-PSDB; AAY06461.
XX
PT   Human proteins for treating and detecting immune and neurological
PT   disorders
XX
PS   Claim 2; Page 110-112; 132pp; English.
XX
CC   This is the nucleotide sequence of cDNA clone HFVIF40, which codes
CC   for human Dendriac (see AAY06461), also called Brainiac-2, a novel
CC   member of the Brainiac family.  Dendriac cDNA is deposited as
CC   ATCC 203056 and ATCC 209627.  The cDNA was initially discovered in
CC   a cDNA library derived from dendritic cells.  Additional clones of
CC   the same gene were identified in NTERA2 cell, adult pulmonary
CC   tissue, salivary gland, ovary, Caco-2 colon adenocarcinoma, smooth
CC   muscle, cerebellum, 8-week-old human mrbyo, haemagiopericytoma,
CC   amygdala, substantia nigra and whole brain cDNA libraries.  Dendriac
CC   nucleic acids are used in the recombinant production of Dendriac
CC   polypeptides, especially the mature protein or epitope-bearing
CC   fragments.  Dendriac and Brainiac proteins are members of the Notch
CC   family of proteins involved in the control of cell division.  The
CC   proteins may be used to detect and/or treat immune system and
CC   neurological disorders.
XX
SQ   Sequence 2189 BP; 657 A; 402 C; 453 G; 677 T; 0 other;

  Query Match             14.0%;  Score 130.4;  DB 20;  Length 2189;
  Best Local Similarity   48.9%;  Pred. No. 1.7e-29;
  Matches  387;  Conservative    0;  Mismatches  396;  Indels    9;  Gaps    1;

Qy    111 TTTACAAGAAAGACGGGAACTTCCTTAAGCTCCCAGATACAGACTGCAGGCAGACACCTC 170
          ||||||   |||||    |||||       |      || || |||||   ||   |  ||
Db    571 TTTACAGACAAGACTTTCACTTCACACTTCGAGAGCATTCAAACTGCTCTCATCAAAATC 630

Qy    171 CCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTTGGCTGAGCGCATGGCCATCC 230
          | || || ||| | ||||||||||||   ||    | |  ||   |   |   ||||||
Db    631 CATTTCTGGTCATTCTGGTGACCTCCCACCCTTCAGATGTGAAAGCCAGGCAGGCCATTA 690

Qy    231 GGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCAGCTGAAGACATTCTTCCTCC 290
          |    || |||||  || | | |     | |||| |  || |    ||||| ||| |
Db    691 GAGTTACTTGGGGTGAAAAAAAGTCTTGGTGGGGATATGAGGTTCTTACATTTTTCTTAT 750

Qy    291 TGGGGACCACCAGCAGTGCAGCGGAAACGAAAGAGG---------TGGACCAGGAGAGCC 341
          | ||                 |  || |  |    ||          | ||  | ||   ||
Db    751 TAGGCCAAGAGGCTGAAAAGGAAGACAAAATGTTGGCATTGTCCTTAGAGGATGAACACC 810

Qy    342 AGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTATTACAATCTGACCCTGA 401
            |   | || ||||| |||    | |||||  |||||   ||| | || ||||||| |||
Db    811 TTCTTTATGGTGACATAATCCGACAAGATTTTTTAGACACATATAATAACCTGACCTTGA 870

Qy    402 AGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCAGGCGGCGTTTGTGATGA 461
          | ||||| ||||   |    |||||   |   |||||  ||  | ||   ||  || ||||
Db    871 AAACCATTATGGCATTCAGGTGGGTAACTGAGTTTTGCCCCAATGCCAAGTACGTAATGA 930

Qy    462 AAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGAACTGCTTCTGAAGAAAA 521
          | |||||| | ||  | |||||||||  || | || |     |    ||| | ||   ||
Db    931 AGACAGACACTGATGTTTTCATCAATACTGGCAATTTAGTGAAGTATCTTTTAAACCTAA 990

Qy    522 ACAGAACAACCAGGTTTTTCACTGGCTTCTTGAAACTCAATGAGTTTCCCATCAGGCAGC 581
          ||    ||   | |||||||||| || |       | |  || | | | ||   ||
Db    991 ACCACTCAGAGAAGTTTTTCACAGGTTATCCTCTAATTGATAATTATTCCTATAGAGGAT 1050

Qy    582 CATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCCGTGGGACAGGTACCCACCATTCT 641
            | |   ||       | |   | |    || ||||| |   |   || ||| |||| ||
Db   1051 TTTACCAAAAAACCCATATTTCTTACCAGGAGTATCCTTTCAAGGTGTTCCCTCCATACT 1110

Qy    642 GCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGCGAGTCAGGTGTACAATGTCTCCA 701
          ||   ||    || ||  |  | ||  | ||  ||| |     | | ||  |  |
Db   1111 GCAGTGGGTTGGGTTATATAATGTCCAGAGATTTGGTGCCAAGGATCTATGAAATGATGG 1170

Qy    702 AGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGTGGGGCTCTGCCTCGAAAGGCTGA 761
            |||   |  ||| ||  | ||||| || | ||| ||| |||| |  |     | |
Db   1171 GTCACGTAAAACCCATCAAGTTTGAAGATGTTTATGTCGGGATCTGTTTGAATTTATTAA 1230

Qy    762 ACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTTTTTTCCAGGGGGCTTACGCTTCT 821
          |  | |   |  |    |   |   |          || |||| |    |  | |  ||
Db   1231 AAGTGAACATTCATATTCCAGAAGACACAAATCTTTTCTTTCTATATAGAATCCATTTGG 1290

Qy    822 CCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTTCATCAAGCCTCGGACTCTCTTGG 881
          || || |   | ||  |  |  | ||   |||   | |     |   |    || |
Db   1291 ATGTCTGTCAACTGAGACGTGTGATTGCAGCCCATGGCTTTTCTTCCAAGGAGATCATCA 1350

Qy    882 ACTACTGGCAGG 893
            |  |||||||
```

```
Db   1351 CTTTTTGGCAGG 1362


RESULT 13
AAX35710
ID   AAX35710 standard; cDNA; 1266 BP.
XX
AC   AAX35710;
XX
DT   09-JUL-1999  (first entry)
XX
DE   cDNA encoding a protein identified by the signal sequence trap method.
XX
KW   Signal sequence trap method; SST method; immunisation; inhibition;
KW   infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW   activin activity; inhibin activity; chemokine activity;
KW   cytokine activity; blood coagulation regulation; agonist; antagonist;
KW   metabolic disorder; hormonal disorder; immune disorder;
KW   severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
KW   wound; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
XX
DR   WPI; 1999-277254/23.
DR   P-PSDB; AAY02366.
XX
PT   Polypeptides identified by the signal sequence trap method from a
PT   human cDNA library
XX
PS   Claim 4; Page 108-109; 281pp; Japanese.
XX
CC   AAX35694-X35747 represent cDNA sequences that encode novel polypeptides
CC   (AAY02358-84) which are identified from a human placental cDNA library
CC   by the signal sequence trap (SST) method. The polypeptides have a
CC   broad range of physiological activity, including immunisation against
CC   and inhibition of infections, allergies and cancer; regulation of tissue
CC   formation and repair; activin/inhibin activity; chemokine/cytokine
CC   activity; blood coagulation regulation; and receptor/ligand agonist
CC   or antagonist activity. The polypeptides can be used for prevention
CC   and treatment of disorders including infections by bacteria, yeasts and
CC   viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC   immune disorders (including severe combined immunodeficiency (SCID)
CC   and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX
SQ   Sequence 1266 BP; 408 A; 270 C; 242 G; 346 T; 0 other;

  Query Match             13.1%;  Score 122.4;  DB 20;  Length 1266;
  Best Local Similarity   50.4%;  Pred. No. 3.7e-27;
  Matches  384;  Conservative    0;  Mismatches  366;  Indels   12;  Gaps    3;

Qy    151 AGACTGCAGGCAGACACCTCCCTTCCTCGTCCTGCTGGTGACCTCATCCCACAAACAGTT 210
Db    432 AAAATGCCAAGAGAAAAGTCCTTTTTTAATACTACTAATAGCTGCAGAGCCTGGACAAAT 491

Qy    211 GGCTGAGCGCATGGCCATCCGGCAGACGTGGGGGAAAGAGAGGATGGTGAAGGGAAAGCA 270
Db    492 AGAAGCTAGAAGAGCTATTCGGCAAACTTGGGGCAATGAAAGTCTAGCACCTGGTATTCA 551

Qy    271 GCTGAAGACATTCTTCCTCCTGGGGACCACCA-GCAGTGCAGCGGAAACGAAAGAGGTGG 329
Db    552 AATCACAAGAATATTTTTGTTGGGCTTAAGTATTAAGCTAAATGGCTACCTTCAACGTGC 611

Qy    330 A-----CCAGGAGAGCCAGCGACACGGGGACATTATCCAGAAGGATTTCCTAGACGTCTA 384
Db    612 AATACTGGAAGAAAGCAGACAATATCATGATATAATTCAACAGGAATACTTAGATACGTA 671

Qy    385 TTACAATCTGACCCTGAAGACCATGATGGGCATAGAATGGGTCCATCGCTTTTGTCCTCA 444
Db    672 CTATAATTTGACCATTAAAACACTAATGGGCATGAACTGGGTTGCAACATACTGTCCACA 731

Qy    445 GGCGGCGTTTGTGATGAAAACAGACTCAGACATGTTCATCAATGTTGACTATCTGACTGA 504
Db    732 TATTCCATATGTTATGAAAACTGACAGTGACATGTTTGTCAACACTGAATATTTAATCAA 791

Qy    505 ACTGCTTCTGAAGAAAAACAGAACAACCAG------GTTTTTCACTGGCTTCTTGAAACT 558
Db    792 TAAGTTACTGAAGCCAGATCTGCCTCCCAGACATAACTATTTCACTGGTTACCTAATGCG 851

Qy    559 CAATGAGTTTCCCATCAGGCAGCCATTCAGCAAGTGGTTTGTCAGTAAATCTGAATATCC 618
Db    852 AGGATATGCACCCAATCGAAACAAAGATAGCAAGTGGTACATGCCACCAGACCTCTACCC 911

Qy    619 GTGGGACAGGTACCCACCATTCTGCTCCGGCACCGGCTACGTGTTTTCTGGCGACGTGGC 678
Db    912 AAGTGAGCGTTATCCTGTCTTCTGTTCTGGAACTGGTTATGTTTTTTCTGGAGATCTGGC 971

Qy    679 GAGTCAGGTGTACAATGTCTCCAAGAGCGTCCCATACATTAAACTGGAAGACGTGTTTGT 738
Db    972 AGAAAAGATTTTTAAAGTTTCTTTAGGTATCCGCCGTTTGCACTTGGAAGATGTATATGT 1031

Qy    739 GGGGCTCTGCCTCGAAAGGCTGAACATCAGATTGGAGGAGCTCCACTCCCAGCCGACCTT 798
Db   1032 AGGGATCTGTCTTGCCAAGTTGAGAATTGATCCTGTACCCCCTCCCAATGAGTTTGTGTT 1091

Qy    799 TTTTCCAGGGGGCTTACGCTTCTCCGTATGCCTCTTCAGGAGGATCGTGGCCTGCCACTT 858
Db   1092 CAATCACTGGCGAGTCTCTTATTCGAGCTGTAAATACAGCCACCTAATTACCTCTCATCA 1151

Qy    859 CATCAAGCCTCGGACTCTCTTGGACTACTGGCAGGCTCTAGA 900
Db   1152 GTTCCAGCCTAGTGAACTGATAAAATACTGGAACCATTTACA 1193


RESULT 14
AAX35711
ID   AAX35711 standard; cDNA; 2420 BP.
XX
AC   AAX35711;
XX
DT   09-JUL-1999  (first entry)
XX
DE   cDNA encoding a protein identified by the signal sequence trap method.
XX
KW   Signal sequence trap method; SST method; immunisation; inhibition;
KW   infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW   activin activity; inhibin activity; chemokine activity;
KW   cytokine activity; blood coagulation regulation; agonist; antagonist;
KW   metabolic disorder; hormonal disorder; immune disorder;
KW   severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
KW   wound; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
```